

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:06:53 ; Search time 170 Seconds
(without alignments)
1758.622 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGIAESKFNKNGKEGSATVLDKNNISKSTNNPK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1dDec04:
1: Geneseqp19808:*
2: Geneseqp19908:*
3: Geneseqp20008:*
4: Geneseqp20018:*
5: Geneseqp20028:*
6: Geneseqp20038:*
7: Geneseqp20038:*
8: Geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	4 AAB48343	Aab48343 S. pneumo
2	4026	100.0	2140	6 ABU01020	Abu01020 S. pneumo
3	4026	100.0	2140	6 ABU45746	Abu45746 Protein e
4	4026	100.0	2140	8 ADM92113	Adm92113 S. pneumo
5	4015.5	99.7	2120	3 AAY81710	Aay81710 Streptoco
6	3789.5	94.1	2138	8 ADK48759	Adk48759 Streptoco
7	3048	75.7	637	8 ADR94534	Adr94534 Streptoco
8	750.5	18.6	1529	8 ADR96136	Adr96136 Novel S.
9	615	15.3	117	2 AAW55096	Aaw55096 Streptoco
10	615	15.3	117	5 ABP54590	Abp54590 S. pneumo
11	615	15.3	117	7 ADC45149	Adc45149 S. pneumo
12	227	5.6	1196	6 ABU24813	Abu24813 Protein e
13	222	5.5	861	7 ABQ23508	Abq23508 Plasmodin
14	220	5.5	1166	8 ADR99186	Adr99186 Streptoco
15	218	5.4	1639	2 AAW54145	Aaw54145 P. falcip
16	218	5.4	1639	5 AAE29345	Aae29345 Plasmodin
17	217	5.4	1166	8 ADR99185	Adr99185 Streptoco
18	217	5.4	1166	8 ADR99178	Adr99178 Streptoco
19	217	5.4	1166	8 ADR99184	Adr99184 Streptoco
20	216	5.4	1166	8 ADR99183	Adr99183 Streptoco
21	216	5.4	1233	5 ABP30203	Abp30203 Streptoco
22	216	5.4	1233	5 ABP29675	Abp29675 Streptoco
23	216	5.4	1233	8 ADR99176	Adr99176 Streptoco
24	216	5.4	1233	8 ADQ10469	Adq10469 Group B S
25	216	5.4	1239	5 ABP25822	Abp25822 Streptoco

26	215.5	5.4	1254	2 AAR07503	Aar07503 Merozite
27	215.5	5.4	1254	2 AAW24575	Aaw24575 Merozite
28	215	5.3	1166	8 ADR99179	Adr99179 Streptoco
29	212.5	5.3	2485	3 AAB18172	Aab18172 Plasmodin
30	212	5.3	1141	6 ABU42327	Abu42327 Protein e
31	212	5.3	1188	3 AAB18183	Aab18183 Plasmodin
32	211.5	5.3	991	8 ADP86452	Adp86452 Clostridi
33	211.5	5.3	1104	7 ADG73658	Adg73658 C. perfir
34	209	5.2	1558	3 AAB18324	Aab18324 Plasmodin
35	208.5	5.2	3029	8 ADP25433	Adp25433 Plasmodin
36	207.5	5.2	4688	6 ABU48941	Abu48941 Protein e
37	207	5.1	1191	6 ABU24124	Abu24124 Protein e
38	207	5.1	1654	1 AAP50777	Aap50777 Sequence
39	207	5.1	1979	3 AAB18171	Aab18171 Plasmodin
40	205	5.1	2024	8 ADP25444	Adp25444 Plasmodin
41	204.5	5.1	1516	3 AAB18195	Aab18195 Plasmodin
42	204.5	5.1	2500	3 AAB18272	Aab18272 Plasmodin
43	204	5.1	1166	6 AAB18982	Aab18982 Pathogen
44	204	5.1	1245	3 AAB18244	Aab18244 Plasmodin
45	203	5.0	1166	2 AAY08643	Aay08643 S. aureus

ALIGNMENTS

RESULT 1
ID AAB48343 standard; protein; 773 AA.
XX
XX AAB48343;
XX
XX 20-APR-2001 (first entry)
XX
XX S. pneumoniae Spi30 polypeptide.
XX
XX Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;
XX bronchial; lung; blood; infection; immune response; immunotherapy;
XX antibacterial; auditory; vaccine.
XX
XX Streptococcus pneumoniae.
XX
XX WO200076540-A2.
XX
XX 21-DEC-2000.
XX
XX 09-JUN-2000; 2000MO-US015925.
XX
XX 10-JUN-1999; 99US-0138453P.
XX
XX (MED1-) MED IMMUNE INC.
XX
XX Adamou JE, Choi GH;
XX WPI: 2001-112197/12.
XX N-PSDB; AAC84742.
XX
XX New vaccines comprising Spi28 or Spi30 polypeptides, for treating and
XX preventing pneumococcal infections, particularly infections caused by
XX Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
XX blood infections.
XX
XX Claim 8; Page 51-54; 54pp; English.
XX
XX The invention relates to novel immunogenic polypeptides, Spi28 and Spi30
XX from S. pneumoniae. Vaccines comprising the polypeptides are useful for
XX the treatment and prevention of pneumococcal infections, particularly
XX infections caused by Streptococcus, such as otitis media, nasopharyngeal,
XX bronchial, lung or blood infections. The antigens are used as immunogenic
XX agents to stimulate an immune response. The antisera and antibodies may
XX also be used in diagnosing and treating pneumococcal infections.
XX Recombinant polypeptides serve as a mechanism for stimulating production
XX of antibodies for use in passive immunotherapy, diagnostic reagents, and
XX as reagents in other processes such as affinity chromatography. The

CC present sequence represents the S. pneumoniae Sp130 polypeptide
XX
SQ Sequence 773 AA;

Query Match	100.0%	Score 4026;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 7.5e-225;		
Matches 773; Conservative	0;	Mismatches 25;	Indels 0;	Gaps 0;

Qy	1	UJGEIAESFERNLGNCKEGLKQDPTGVHHHOBNEISIRKSSPTIDRNIISTIRDEPNK	60
Dp	1	KGBIAESKPNLGVNGKEGSLKQDPTGVHHHOBNEISIRKSSPTIDRNIISTIRDEPNK	60
Qy	61	DLKCKIKKKFRAVDPTSETGKMBEYDYKDYDKGNIAAYDGDGTLEFYEIRKDEIRSKI	120
Dp	61	DLKCKIKKKFRAVDPTSETGKMBEYDYKDYDKGNIAAYDGDGTLEFYEIRKDEIRSKI	120
Qy	121	YGVLSPSKDGAFBEILGKISVNSRNAFVYGGNNYSIEIKATKYDFHSKTYTFDLYANIND	180
Dp	121	YGVLSPSKDGAFBEILGKISVNSRNAFVYGGNNYSIEIKATKYDFHSKTYTFDLYANIND	180
Qy	181	IYDGLAFAGDMLPYKDNDQKAEIICIRPEKIKETKSEBPYVYSYGNNIIELEBGDLSKN	240
Dp	181	IYDGLAFAGDMLPYKDNDQKAEIICIRPEKIKETKSEBPYVYSYGNNIIELEBGDLSKN	240
Qy	241	KPDNLTKMBSGKLYSDSEKQOYLTKONIILRKGYALKVTTYNGKJDMLEGNVYSKEDI	300
Dp	241	KPDNLTKMBSGKLYSDSEKQOYLTKONIILRKGYALKVTTYNGKJDMLEGNVYSKEDI	300
Qy	301	AKIQANPNLALSETTYIYASRNVEDGKSTOSVLSALDGFNIIRYOVTFKQNDGKA	360
Dp	301	AKIQANPNLALSETTYIYASRNVEDGKSTOSVLSALDGFNIIRYOVTFKQNDGKA	360
Qy	361	IDKQGNLYTDSKLYLFGDDKXYTGEDKPNYKALIKEDSGMLFTDTPKPNVLSMDKRYENP	420
Dp	361	IDKQGNLYTDSKLYLFGDDKXYTGEDKPNYKALIKEDSGMLFTDTPKPNVLSMDKRYENP	420
Qy	421	SKSNKIIYYRNPBYLKGKISDKGSPNMBIRKVNESVDNPLYIYEDLHIDNTRDPIIKLNVK	480
Dp	421	SKSNKIIYYRNPBYLKGKISDKGSPNMBIRKVNESVDNPLYIYEDLHIDNTRDPIIKLNVK	480
Qy	481	DGDIIMDGKMDKYKANGPDPKVTIDMDGNVYLQNGYSDLNKAQVGHYQFLYDNYKPEVNIID	540
Dp	481	DGDIIMDGKMDKYKANGPDPKVTIDMDGNVYLQNGYSDLNKAQVGHYQFLYDNYKPEVNIID	540
Qy	541	PKGNITSIYADOKSYVFNINDKRNNGFDEIRIOBHUYINGKRYTSPNDIKQIIDKTLNIK	600
Dp	541	PKGNITSIYADOKSYVFNINDKRNNGFDEIRIOBHUYINGKRYTSPNDIKQIIDKTLNIK	600
Qy	601	IYVNDPANTTYKBEPLINDNGEVSLEIKRHYVYTIQNKREMSSTIYSEBDFLLPYVKGK	660
Dp	601	IYVNDPANTTYKBEPLINDNGEVSLEIKRHYVYTIQNKREMSSTIYSEBDFLLPYVKGK	660
Qy	661	LEKGYQPDGWEISGFEKGKADAYVNLISDTEFLKPPYKLIKBEKGESEBNKTPYVSKKKN	720
Dp	661	LEKGYQPDGWEISGFEKGKADAYVNLISDTEFLKPPYKLIKBEKGESEBNKTPYVSKKKN	720
Qy	721	POVNHQSQLESIRKEDLOREHSQKSDSYEDVYATVYTLNDNNISSKSTNNPNK 773	
Dp	721	POVNHQSQLESIRKEDLOREHSQKSDSYEDVYATVYTLNDNNISSKSTNNPNK 773	

RESULT 2	
ABU01020	
ID	ABU01020 standard; protein; 2140 AA.
XX	
AC	ABU01020;
XX	
DT	23-OCT-2003 (revised)
DT	11-FEB-2003 (first entry)
XX	
DE	S. pneumoniae type 4 strain protein from coding region #590.
XX	
XX	Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection

KM antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KM gene therapy; vaccine.

05 *Streptococcus pneumoniae*; type 4 strain.


PN WO200277021-A2

PD 03-OCT-2002.

PF 27-MAR-2002; 2002WO-IB002163.

PR 27-MAR-2001; 2001GB-00007658.

PA (CHIR-) CHIRON SPA.



XX

DR N-PSDB; ABX06302.

PT New proteins and

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or ear infection.

PS Claim 1; SEQ ID NO 1180; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid residues, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as AB556454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus* nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a *Streptococcus pneumoniae* bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences). (updated on 23-OCT-2003 to standardise .ins field)

SQ Sequence 2140 AA;

Query Match	100.0%	Score 4026;	DB 6;	Length 2140;
Best Local Similarity	100.0%	Pred. No. 2.9e-224;		
Matches 773; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	KUGJAEKFNOLANGKRGSLKOTTCYBHHHONNESSIKKSSFTIDRNISTIDFENK	60
Db	1334	KUGJAEKFNOLANGKRGSLKOTTCYBHHHONNESSIKKSSFTIDRNISTIDFENK	1393
QY	61	DLKGLIKKKFRRVDDFTSETKRMEEYDYKTDYDKGNIIAYYDGTDLBYETEKLDKIKSI	120
Db	1394	DLKGLIKKKFRRVDDFTSETKRMEEYDYKTDYDKGNIIAYYDGTDLBYETEKLDKIKSI	1455

```

QY 121 YGVLSPSKDGHEILGKISNYSKNAKVVYGNVYKSIIEIKATKYDPSKMTEDLVANIND 180
DB 1454 YGVLSPSKDGHEILGKISNYSKNAKVVYGNVYKSIIEIKATKYDPSKMTEDLVANIND 1513
QY 181 IVDGLAFAGDMRLFYKNDNQKKAIRIKRPERIKETKSEFPYVSSYGVNILEGEGDLSKN 240
DB 1514 IVDGLAFAGDMRLFYKNDNQKKAIRIKRPERIKETKSEFPYVSSYGVNILEGEGDLSKN 1573
QY 241 KPDNLTWESGKITYSDSEKQOYLAKNDIILRGYALKVTTNPGKTDMLEGNGVYSKEKI 300
DB 1574 KPDNLTWESGKITYSDSEKQOYLAKNDIILRGYALKVTTNPGKTDMLEGNGVYSKEKI 1633
QY 301 AKIQKAMPNLRALSETTIYADSRNVEDGSRSTQSVMSALDGNRIIRYQVFTFRMNDKGBA 360
DB 1634 AKIQKAMPNLRALSETTIYADSRNVEDGSRSTQSVMSALDGNRIIRYQVFTFRMNDKGBA 1693
QY 361 IDKGNLVTDSKLVLPFGKDKKXTYGBDKFNVAIKEDGSMFLPDTKPVNLSMDKNYENP 420
DB 1694 IDKGNLVTDSKLVLPFGKDKKXTYGBDKFNVAIKEDGSMFLPDTKPVNLSMDKNYENP 1753
QY 421 SKSNKTIYVRNPEFYLKRGKISDKGFGNMBLKVNESVVDNVLTYGDLHIDNTRDPIKLVNK 480
DB 1754 SKSNKTIYVRNPEFYLKRGKISDKGFGNMBLKVNESVVDNVLTYGDLHIDNTRDPIKLVNK 1813
QY 481 DGDIMDWGKDYKANGFPDKVTDMDGNVYLQTVSGDLNAKAVGVHYQFLYDNVKEBWNID 540
DB 1814 DGDIMDWGKDYKANGFPDKVTDMDGNVYLQTVSGDLNAKAVGVHYQFLYDNVKEBWNID 1873
QY 541 PKGNTSITYADGKSVVENINDKRNNGFDEIOBOHIYNGKETSFPNDIKQIIDXTLVNK 600
DB 1874 PKGNTSITYADGKSVVENINDKRNNGFDEIOBOHIYNGKETSFPNDIKQIIDXTLVNK 1933
QY 601 IVVNDPANTTYKKEPILANDTGEVSELKPHRYVTYITLONKEMSSITVSEEDPILPYVGE 660
DB 1934 IVVNDPANTTYKKEPILANDTGEVSELKPHRYVTYITLONKEMSSITVSEEDPILPYVGE 1993
QY 661 LEKGYPDGMWISGFGKXKADAGVYINLSKDTPIKVPFKIKBEKKBEKNPPTDVSKKDN 720
DB 1994 LEKGYPDGMWISGFGKXKADAGVYINLSKDTPIKVPFKIKBEKKBEKNPPTDVSKKDN 2053
QY 721 PQVNHSQLNBSHRKEDLQREHSQKSDSTKYTATVYLDNNISSTSTTNNPK 773
DB 2054 PQVNHSQLNBSHRKEDLQREHSQKSDSTKYTATVYLDNNISSTSTTNNPK 2106

```

RESULT 3

ABU45746
ID ABU45746 standard; protein; 2140 AA.

```

AC ABU45746;
DB 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #31273.
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS Streptococcus pneumoniae.
XX MO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002MO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (BLIT-) BLITRA PHARM INC.

```

```

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zvekind JW,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
XX WPI; 2003-029926/02.
DR N-PSDB; ACA49616.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
PS Claim 25; SEQ ID NO 73670; 1766bp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 623 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp://wipo.int/pub/published_pct_sequences
XX
XX
XX Sequence 2140 AA:
XX
XX Query Match 100.0%; Score 4026; DB 6; Length 2140;
XX Best Local Similarity 100.0%; Pred. No. 2.9e-224;
XX Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 361 IDKDGMLVTDSKLV.FGKDKKRYTGEDKFNVAIKEDSGMLPIOTFVNLSDMKRYNP 420
 DB 1694 IDKDGMLVTDSKLV.FGKDKKRYTGEDKFNVAIKEDSGMLPIOTFVNLSDMKRYNP 1753
 QY 421 SKSNKIYVNPBEFLYLGKISDKGCFWMLRVNESVVDNYLIYGDHIDNTRDFNIIKLANVK 480
 DB 1754 SKSNKIYVNPBEFLYLGKISDKGCFWMLRVNESVVDNYLIYGDHIDNTRDFNIIKLANVK 1813
 QY 481 DDDIMDMGKDYKANGPPDKVTDMDGNVTLQYGSIDLNAKAVGVHQLYDNVKEPVNID 540
 DB 1814 DDDIMDMGKDYKANGPPDKVTDMDGNVTLQYGSIDLNAKAVGVHQLYDNVKEPVNID 1873
 QY 541 PKGNTSIEYADGKSVFENINDKRNNGPDEIOBOHIYINGKETSFPNDIKOIIDKTLNKK 600
 DB 1874 PKGNTSIEYADGKSVFENINDKRNNGPDEIOBOHIYINGKETSFPNDIKOIIDKTLNKK 1933
 QY 601 IYVKDPARTTYKBEFLINKDTGEVSELKPHRVVTIIONGKEMSTIVSEDFILPYKGE 660
 DB 1934 IYVKDPARTTYKBEFLINKDTGEVSELKPHRVVTIIONGKEMSTIVSEDFILPYKGE 1993
 QY 661 LEKGYPDGWEISGFGKGDAGVYNLSKDTPIKVPFKKIEKKESEBNKPTFVDSKKON 720
 DB 1994 LEKGYPDGWEISGFGKGDAGVYNLSKDTPIKVPFKKIEKKESEBNKPTFVDSKKON 2053
 QY 721 POWNHSQLNESHKEDLQREBSHOKSDSTKDYATATVLDKNNISSKSTNNPNK 773
 DB 2054 POWNHSQLNESHKEDLQREBSHOKSDSTKDYATATVLDKNNISSKSTNNPNK 2106

RESULT 4

ADM92113
 ID ADM92113 standard; protein; 2140 AA.

AC ADM92113;

XX 03-JUN-2004 (first entry)

DE S pneumoniae antigenic protein sequence SeqIDJ10.

KW antibacterial; gene therapy; Streptococcus pneumoniae infection;

KM antigenic.

OS Streptococcus pneumoniae.

PN WO2004020609-A2.

XX 11-MAR-2004.

PD 02-SBP-2003; 2003WO-US027401.

XX 30-AUG-2002; 2002US-0407082P.

PR (TUPT) UNIV TUPITS.

XX Camilla A, Hava DL;

XX WPI; 2004-239189/22.

DR N-PSDB; ADM91876.

XX New Streptococcus pneumoniae nucleic acid molecules, useful for
 PT diagnosing, treating and preventing active infections of Streptococcus
 PT pneumoniae.

PS Claim 27; SEQ ID NO 310; 123pp; English.

XX This invention relates to novel isolated Streptococcus pneumoniae nucleic
 CC acid molecules and the antigenic polypeptides encoded by them. The
 CC invention may be useful for the production of compounds with an
 CC antibacterial activity or for gene therapy. The nucleic acid molecules,
 CC compositions and methods disclosed are useful for treating Streptococcus
 CC pneumoniae infection. The present sequence is that of an S pneumoniae
 CC protein of the invention.

XX SQ Sequence 2140 AA;

Query Match 100.0%; Score 4026; DB 8; Length 2140;
 Best Local Similarity 100.0%; Pred. No. 2,9e-224;
 Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLEFIASRYKRNIGNGKESGLKQTTGVHEHHOENBSIYKSSPTIDRNIISTIRPENK 60
 DB 1334 KLEFIASRYKRNIGNGKESGLKQTTGVHEHHOENBSIYKSSPTIDRNIISTIRPENK 1393
 QY 61 DLKGLIKKREAVDDFTSFGKRMERYDYKYDDKGNIIAODGPDLEFETKLDERTSKI 120
 DB 1394 DLKGLIKKREAVDDFTSFGKRMERYDYKYDDKGNIIAODGPDLEFETKLDERTSKI 1453
 QY 121 YGVLSPSKGFHFIILKISNVSNARVYVYGNVSKSIEIKATKYDFHSKTWTFDLNANIND 180
 DB 1454 YGVLSPSKGFHFIILKISNVSNARVYVYGNVSKSIEIKATKYDFHSKTWTFDLNANIND 1513
 QY 181 IYVGLAPADGMLFVNDNQKAEIKLIMPEKIKETKSEVPYSSYGNVIELGEGDLSKN 240
 DB 1514 IYVGLAPADGMLFVNDNQKAEIKLIMPEKIKETKSEVPYSSYGNVIELGEGDLSKN 1573
 QY 241 KPDMLTMSGKIYSDSKQOYLKONIIILKGYALKVTYVNGKTDMLGNGVYSKEDI 300
 DB 1574 KPDMLTMSGKIYSDSKQOYLKONIIILKGYALKVTYVNGKTDMLGNGVYSKEDI 1633
 QY 301 AKIQKANPMLRALSETTIYADSHNVEDGRTOSVLSALDGFNIIRYQVETFMNDKGBA 360
 DB 1634 AKIQKANPMLRALSETTIYADSHNVEDGRTOSVLSALDGFNIIRYQVETFMNDKGBA 1693
 QY 361 IDKDGMLVTDSKLV.FGKDKKRYTGEDKFNVAIKEDSGMLPIOTFVNLSDMKRYNP 420
 DB 1694 IDKDGMLVTDSKLV.FGKDKKRYTGEDKFNVAIKEDSGMLPIOTFVNLSDMKRYNP 1753
 QY 421 SKSNKIYVNPBEFLYLGKISDKGCFWMLRVNESVVDNYLIYGDHIDNTRDFNIIKLANVK 480
 DB 1754 SKSNKIYVNPBEFLYLGKISDKGCFWMLRVNESVVDNYLIYGDHIDNTRDFNIIKLANVK 1813
 QY 481 DDDIMDMGKDYKANGPPDKVTDMDGNVTLQYGSIDLNAKAVGVHQLYDNVKEPVNID 540
 DB 1814 DDDIMDMGKDYKANGPPDKVTDMDGNVTLQYGSIDLNAKAVGVHQLYDNVKEPVNID 1873
 QY 541 PKGNTSIEYADGKSVFENINDKRNNGPDEIOBOHIYINGKETSFPNDIKOIIDKTLNKK 600
 DB 1874 PKGNTSIEYADGKSVFENINDKRNNGPDEIOBOHIYINGKETSFPNDIKOIIDKTLNKK 1933
 QY 601 IYVKDPARTTYKBEFLINKDTGEVSELKPHRVVTIIONGKEMSTIVSEDFILPYKGE 660
 DB 1934 IYVKDPARTTYKBEFLINKDTGEVSELKPHRVVTIIONGKEMSTIVSEDFILPYKGE 1993
 QY 661 LEKGYPDGWEISGFGKGDAGVYNLSKDTPIKVPFKKIEKKESEBNKPTFVDSKKON 720
 DB 1994 LEKGYPDGWEISGFGKGDAGVYNLSKDTPIKVPFKKIEKKESEBNKPTFVDSKKON 2053
 QY 721 POWNHSQLNESHKEDLQREBSHOKSDSTKDYATATVLDKNNISSKSTNNPNK 773
 DB 2054 POWNHSQLNESHKEDLQREBSHOKSDSTKDYATATVLDKNNISSKSTNNPNK 2106

RESULT 5

AA81710
 ID AA81710 standard; protein; 2120 AA.

XX AA81710;

XX 02-JUN-2000 (first entry)

DB Streptococcus pneumoniae protein sequence ID3.

XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;

KW pneumococcal septicemia; sinusitis; meningitis; therapy.
 XX Streptococcus pneumoniae.
 OS WO200006738-A2.
 FN 10-FEB-2000.
 PD 27-JUL-1999; 99WO-GB002452.
 XX 27-JUL-1998; 98GB-00016336.
 PR 19-MAR-1999; 99US-0125329P.
 XX (MICR-) MICROBIAL TECHNIQS LTD.
 PA
 XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
 PI WPI; 2000-199301/17.
 DR N-PSDB; AA291806.
 XX Streptococcal proteins and polynucleotides useful for diagnosis,
 PT treatment and prophylaxis of bacterial infections.
 XX
 XX Claim 2; Page 41-42; 76pp; English.
 CC This sequence represents a Streptococcus pneumoniae protein of the
 CC invention. The proteins (or their homologues, derivatives and/or
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
 CC compositions comprising the proteins are useful as vaccines and also in
 CC diagnostic assays. The sequences are useful for the detection or
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
 CC with them. Agents capable of antagonising, inhibiting or interfering with
 CC the function or expression of the protein or polypeptide are useful in
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae
 CC infection. As the sequences can be used to treat S. pneumoniae infection,
 CC they can be used to treat bacterial pneumonia, which has high rates in
 CC young children, the elderly, and in patients with predisposing conditions
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
 CC or with immunosuppressive disorders, especially AIDS. They can also be
 CC used to treat pneumococcal septicemia, otitis media, sinusitis, and
 CC meningitis
 XX
 SO Sequence 2120 AA;
 Query Match 99.7%; Score 4015.5; DB 3; Length 2120;
 Best Local Similarity 99.9%; Pred. No. 1.1e-223;
 Matches 773; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 1673 IDKGNLVYDSSGLVLFKGDKEYTGEDKFNVEALIKEDSMLFIDTKPVNLSDKXVFN 1732
 QY 421 SKSNKTYANPEPYLAGKISDGGFWMLRVNVSVDNLTLYGDLHIDNR-DENIKLV 479
 DB 1733 SKSNKTYANPEPYLAGKISDGGFWMLRVNVSVDNLTLYGDLHIDNRDIPKLV 1792
 QY 480 KGDIDMDGMDKYKANGPPDKYTDMDGNVYLQGYSDLNAKAVGHYQFLYDNVKEVNI 539
 DB 1793 KGDIDMDGMDKYKANGPPDKYTDMDGNVYLQGYSDLNAKAVGHYQFLYDNVKEVNI 1852
 QY 540 DPKGNTSIRYADGKSVYFNINDRNNGFPGGEIOEGHIIYNGKXTSPNDIKQIIDKTLNI 599
 DB 1853 DPKGNTSIRYADGKSVYFNINDRNNGFPGGEIOEGHIIYNGKXTSPNDIKQIIDKTLNI 1912
 QY 600 KTVVQDPAFTWYKKEITLKNDDGEVSELPHRTVYTIQNGKMSSTIVSEEDILPYK 659
 DB 1913 KTVVQDPAFTWYKKEITLKNDDGEVSELPHRTVYTIQNGKMSSTIVSEEDILPYK 1972
 QY 660 ELEKGYQFDGWBISGFEKKGADGVYINLSKDTPIKEVFKIIEKKEENKPTDVSXGD 719
 DB 1973 ELEKGYQFDGWBISGFEKKGADGVYINLSKDTPIKEVFKIIEKKEENKPTDVSXGD 2032
 QY 720 NPQVHNSQANESHKREDLQREHNSQKSDSTKVYATVLDKNNISSTTNPNK 773
 DB 2033 NPQVHNSQANESHKREDLQREHNSQKSDSTKVYATVLDKNNISSTTNPNK 2086
 RESULT 6
 ADK48759
 ID ADK48759 standard; protein; 2138 AA.
 XX
 AC ADK48759;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DB Streptococcus pneumoniae protein, Seq ID No 5274.
 XX
 XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
 OS Streptococcus pneumoniae.
 XX
 PN US6699703-B1.
 XX
 PD 02-MAR-2004.
 XX
 PF 26-MAY-2000; 2000US-00583110.
 XX
 PR 02-JUL-1997; 97US-0051553P.
 PR 12-MAY-1998; 98US-0085131P.
 PR 30-JUN-1998; 98US-00107433.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CB;
 XX
 XX N-PSDB; ADK46098.
 DR
 PT New nucleic acid molecules and polypeptides useful for diagnosing,
 PT preventing and treating pathological conditions resulting from bacterial
 PT infection, e.g. Streptococcus pneumoniae infection, and in drug
 PT screening.
 XX
 PS Disclosure; SEQ ID NO 5274; 301pp; English.
 XX
 CC The invention relates to isolated Streptococcus pneumoniae nucleic acids
 CC and polypeptides. The nucleic acids and proteins are useful for
 CC diagnosing, preventing and treating pathological conditions resulting
 CC from bacterial infection, such as S. pneumoniae infection. These may also
 CC be used for drug screening procedures. The present sequence represents a
 CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
 CC data for this patent did not appear in the printed specification but was

CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.

bequaata.uspto.gov/bequaata.html.

SQ Sequence 2138 AA;

Query Match	94.1%	Score 3789.5	DB 8	Length 2138
PostgreSQL	94.8%	Score 3789.5	DB 8 <td>Length 2138</td>	Length 2138

Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;

Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;

QY	1	KLGEIAEKFNILGNGKESGLKCKDTYTGVEHHQNEBESIKESFTIDIRNITIDFENK	60
Db	1333	KLGEIPEKPEFNILKKNVXDLSLKEIFARFENNLVDNQSIEGSLFIRIKHTISTIDFENK	1392
QY	61	DLKCLIKKKPFREVDFTSTGQMESEYDYKTDKNNIIAYDGTDLAEYTEKDLSEKSI	120
Db	1393	DLKCLIKKKYKQEDDFVN-GGTRFVBYDYKTDKNNIIAYDGTDLAEYTEKDLSEKSI	1451
QY	1231	YGVLSPSDGHPEILIGKISNVSKNANVYGNNGYSIEIKATYDPSHSTMPDLYANIND	180
Db	1452	YGVLSPSDGHPEILIGKISNVSKNANVYGNNGYSIEIKATYDPSHSTMPDLYANIND	1511
QY	161	IYDGLAFAGDMRLFYKNDQKKABIKIIMPEKIKETKSEYPVYSSGVNIVIEIGDLSKN	240
Db	1512	IYDGLAFAGDMRLFYKNDQKKABIKIIMPEKIKETKSEYPVYSSGVNIVIEIGDLSKN	1571
QY	241	KPDNLTNMGSGIYEDSEKQOYLKONIIIRKGYALKTTTNPCTDMLBEGNGVSKEDI	300
Db	1572	KPDNLTNMGSGIYEDSEKQOYLKONIIIRKGYALKTTTNPCTDMLBEGNGVSKEDI	1631
QY	301	AKIQKANENLALSETTIYADSRNVEDGRSTQSYAMSALDGFNIIIRYOVTFPKNDKBA	360
Db	1632	AKIQKANENLALSETTIYADSRNVEDGRSTQSYAMSALDGFNIIIRYOVTFPKNDKBA	1691
QY	361	IDKQGNLVTDSKVLFGKDKERYTGBDKXFNVEALKEQDSMLFIDTRKPNLSMDKNYFP	420
Db	1692	IDKQGNLVTDSKVLFGKDKERYTGBDKXFNVEALKEQDSMLFIDTRKPNLSMDKNYFP	1751
QY	421	SKSNKIIYRANBEFYLRGKISDKGSGNMLARVNESVVDNYLIYGDLEHIDNTDFNKLNVK	480
Db	1752	SKSNKIIYRANBEFYLRGKISDKGSGNMLARVNESVVDNYLIYGDLEHIDNTDFNKLNVK	1811
QY	481	DGDIIMDWGMDKYKANGPPDKYTLDMDGNYVLTQYISDLNAKAVGVHYQFLYDNVPEVINI	540
Db	1812	DGDIIMDWGMDKYKANGPPDKYTLDMDGNYVLTQYISDLNAKAVGVHYQFLYDNVPEVINI	1871
QY	541	PKGNTSIFYAAGKSVYFVINIDKRNNGPFGELIOBHIIYNGKSYTSFNDIKOIIDTLANIK	600
Db	1972	PKGNTSIFYAAGKSVYFVINIDKRNNGPFGELIOBHIIYNGKSYTSFNDIKOIIDTLANIK	1931
QY	601	IYVNDPANTTYKEPIILNKDTGSEVSELKPHVYVYTIQNGKMSSTIVSEEDFILPVYKGE	660
Db	1932	IYVNDPANTTYKEPIILNKDTGSEVSELKPHVYVYTIQNGKMSSTIVSEEDFILPVYKGE	1991
QY	661	LEKGYQDPGMEISGPEGKQDAGYINLSKOTPIKVPFKCIEKKCBENKPTFDVSKKQDN	720
Db	1992	LEKGYQDPGMEISGPEGKQDAGYINLSKOTPIKVPFKCIEKKCBENKPTFDVSKKQDN	2051
QY	721	PQVNHSQLNBSHRKEDLQREBSQKSDSTQVTAIVLADKNNISSKSTNNPNK 773	
Db	2052	PQVNHSQLNBSHRKEDLQREBSQKSDSTQVTAIVLADKNNISSKSTNNPNK 2104	
RESULT 7			
ID	ADR94534	standard; protein; 637 AA.	
XX	ADR94534;		
XX	16-DBC-2004	(first entry)	
XX	Novel S. pneumoniae protein sequence, SEQ ID 3169.		
XX	Meningitis; bacteraemia; pneumonia; otitis media; vaccine;		

KM	bacterial infection.
XX	
OS	<i>Streptococcus pneumoniae</i> .
PN	US680074-B1.
XX	
PD	05-OCT-2004.
XX	
PF	30-JUN-1998; 98US-00107433.
XX	
PR	02-JUL-1997; 97US-0051553P. 12-MAY-1998; 98US-0085131P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
P1	Doucette-Stamm LA, Bush D;
PI	
DR	WPI; 2004-697205/68.
DR	N-PsDB; ADR91931.
XX	
PT	New isolated nucleic acid encoding a <i>Streptococcus pneumoniae</i>
PT	polypeptide, useful for diagnosing, preventing and/or treating
PT	pathological conditions resulting from the bacterial infection
PS	Disclosure; SEQ ID NO 3169; 151bp; English.
XX	

The invention relates to an isolated nucleic acid comprising a sequence encoding a Streptococcus pneumoniae ADR9336polypeptide, or its fragments, with any of 9 fully defined sequences (appearing as ADR943308, ADR944489, ADR944800, ADR944837, ADR949659, ADR95253, ADR95647, ADR95682, ADR96079) or any of the fully defined sequences appearing as ADR911705, ADR91886, ADR921217, ADR92234, ADR93039, ADR93079, ADR92336, ADR92650 or ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide sequence, or at least 40, 60 or 300 consecutive nucleotides, which is hybridisable under high stringency conditions to the nucleotide sequence. The nucleic acids and proteins are chosen from 5206 disclosed sequences. Also included are a recombinant expression vector comprising the isolated nucleic acid cited above operably linked to a transcription regulatory element, a cell comprising the recombinant expression vector and a probe comprising at least 20 consecutive nucleotides of the nucleotide sequence as cited above. The methods and compositions of the present invention are useful for the diagnosis, prevention and/or treatment of pathological conditions resulting from bacterial infection by *Streptococcus pneumoniae* e.g. pneumonia, bacteraemia, meningitis and otitis media. The present sequence is one of the 2603 disclosed 5. CC pneumonias protein sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?docid=6600744B1.

SQ Sequence 637 AA;

Query Match	75.74	Score 3048	DB 8	Length 637
Best Local Similarity	99.34	Pred. NO. 2.8e-168		
Matches 585	Conservative 3	Mismatches 1	Indels 0	Gaps 0

Qy	185	IAFAQGMRLFYVDNQOKRAEIKIRNPEKIKETKSYSPVSSYGAVITELGSEDLSKNRPN	244
Db	15	IAFAQGMRLFYVDNQOKRAEIKIRNPEKIKETKSYSPVSSYGAVITELGSEDLSKNRPN	74
Qy	245	LTPMBSGKIYDSBRQOYLAKNIIILRGVALKYTYNPGKTDMLBNGGVSKEDIAKIQ	304
Db	75	LTPMBSGKIYDSBRQOYLAKNIIILRGVALKYTYNPGKTDMLBNGGVSKEDIAKIQ	134
Qy	305	KANPNIRALSETTIYADSRNVEDGHSSTQSLMSALDGNIIIRYQVPTFRANDKEALDKD	364
Db	135	KANPNIRALSETTIYADSRNVEDGHSSTQSLMSALDGNIIIRYQVPTFRANDKEALDKD	194
Qy	365	GNLVTYDSSKLVLFGKDDKEKTYGEDFTNVEALKEDGSMFLFIDTKRVNLSMDKNYFNPESKN	424
Db	195	GNLVTYDSSKLVLFGKDDKEKTYGEDGDSNVEALKEDGSMFLFIDTKRVNLSMDKNYFNPESKN	254
Qy	425	KIYVRNPEFTYLGKISDKGQFMELRVNBSVVDNYTLIYGDLAHIDNTRPNIKLNVXDGI	484

```

DB 255 KTYANPFFYLGKSIDKSGFHWELRVNNSVDNVIYTGDIHIDNTRDNRIKLNVDGDI 314
QY 485 MDWGNKDYKANGFPDPKVTMDGANNVYIQTSYSDLNAAKAVGVRQFLYDNYKPEVNIIDPKCN 544
DB 315 MDWGNKDYKANGFPDPKVTMDGANNVYIQTSYSDLNAAKAVGVRQFLYDNYKPEVNIIDPKCN 374
QY 545 TSIEFADKSVNPNNDKRNNGPDGEIOGHOIYINKEKYSFNDIQIIDKTNITIVVK 604
DB 375 TSIEFADKSVNPNNDKRNNGPDGEIOGHOIYINKEKYSFNDIQIIDKTNITIVVK 434
QY 605 DPARNTVYKFFILNKDTGSEVSLKPRVYVYTIQNGKMSSTIYSEEDFLIPYKGLKLG 664
DB 435 DPARNTVYKFFILNKDTGSEVSLKPRVYVYTIQNGKMSSTIYSEEDFLIPYKGLKLG 494
QY 665 YQPDGMEISGFEGKCDAGTVINLSKDTPIKPVPKLIEKKKEENKPTPVSKXKDPQVN 724
DB 495 YQPDGMEISGFEGKCDAGTVINLSKDTPIKPVPKLIEKKKEENKPTPVSKXKDPQVN 554
QY 725 HSQLESFHKEDLQREBSQKSDSTQDVATVLDKNISSTKSTNNPNK 773
DB 555 HSQLESFHKEDLQREBSQKSDSTQDVATVLDKNISSTKSTNNPNK 603

```

RESULT 8
ADR96136
ID ADR96136 standard; protein; 1529 AA.

AC ADR96136;
DT 16-DEC-2004 (first entry)

DE Novel S. pneumoniae protein sequence, SEQ ID 4771.

KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
bacterial infection.

OS Streptococcus pneumoniae.

PN US6800744-B1.

PD 05-OCT-2004.

PF 30-JUN-1998; 98US-00107433.

PR 02-JUL-1997; 97US-0051553P.

PR 12-MAY-1998; 98US-0085131P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

DR WPI; 2004-697205/68.

DR N-PSDB; ADR93533.

PT New isolated nucleic acid encoding a Streptococcus pneumoniae
polypeptide, useful for diagnosing, preventing and/or treating
pathological conditions resulting from the bacterial infection.

PS Disclousure; SEQ ID NO 4771; 151pp; English.

CC The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridisable under high stringency conditions to the nucleotide sequence.
CC The nucleic acid and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory

CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae protein sequences. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.

SQ Sequence 1529 AA;

Query Match 18.6%; Score 750.5; DB 8; Length 1529;
Best local Similarity 79.7%; Pred. No. 1,3e-34;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

```

QY 1 KLGRIASRFRKNI/GNGEGSLKKDITGVHGHQENESIKESPTIDRNISTRDFENK 60
DB 1336 KLGRIEBSKFKNLKXKVDLSLNGKTAIVENNLLVDQSLBGSLSFVHHTKISTIRDFENK 1395
QY 61 DLKGLIKKKFREVDPFTSETGKRMERYDYKYDDKNI IAYDGTDLRYETKLDKISKI 120
DB 1396 DLKGLIKKKYKQEDDFVN-GGTRVARDYKYDDKNI IAYDGTDLRYETKLDKISKI 1454
QY 121 YGVLSPSKDGHPRIILAKISNVSKNAKVYGNKYSLIETATKYDFHSKTMDPLVYANIND 180
DB 1455 YGVLSPSKDGHPRIILAKISNVSKNAKVYGNKYSLIETATKYDFHSKTMDPLVYANIND 1514
QY 181 IVDGLAF 187
DB 1515 IVDGISF 1521

```

RESULT 9

AAW55096
ID AAW55096 standard; protein; 117 AA.

AC AAW55096;

DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae SP0043 protein.

KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
detection; pneumonia; otitis media; meningitis.

OS Streptococcus pneumoniae.

PN WO9818930-A2.

PD 07-MAY-1998.

PF 30-OCT-1997; 97WO-US019422.

PR 31-OCT-1996; 96US-0023960P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Kunach CA, Choi GH, Johnson LS, Hiromocky] A;

DR WPI; 1998-272224/24.

DR N-PSDB; AAW7357.

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
or their epitope-containing fragments, useful in protective or
therapeutic vaccines, and for diagnosis.

PS Claim 11; Page 62; 118pp; English.

CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein

CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly for detecting specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose

CC pneumoniae antigens have antibacterial activity and can be used in
CC vaccines. The S. pneumoniae antigens can also be used to prevent or
CC attenuate a Streptococcal infection in an animal. The polynucleotides
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
CC nucleic acids. AB084905 to AB085130 represent primers used in the cloning
CC of S. pneumoniae ORFs (open reading frames) which are used in an example
CC from the present invention

XX Sequence 117 AA;

Query Match 15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.2e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 657 YKGELEKGYQPDGMEISGFGKKDAGVYINLSKDTFKKPKFKKEKKEENKPTDVSX 716
DB 1 YKGELEKGYQPDGMEISGFGKKDAGVYINLSKDTFKKPKFKKEKKEENKPTDVSX 60

OY 717 KKNPQVNHSQLNESHKEDLQREHRSQKSDSTKDYATATVLDKNNTSSKSTNNPK 773
DB 61 KKNPQVNHSQLNESHKEDLQREHRSQKSDSTKDYATATVLDKNNTSSKSTNNPK 117

RESULT 10
ABP54590
ID ABP54590 standard; protein; 117 AA.

XX ABP54590;

DT 04-SEP-2002 (first entry)

XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.

XX Streptococcus pneumoniae.

XX US2002061545-A1.

XX 23-MAY-2002.

XX 22-JAN-2001; 2001US-00765272.

XX 30-OCT-1997; 97US-00961083.

XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.
XX (DOUG/) DOUGHERTY B.
XX (FANN/) FANNON M R.
XX (ROSE/) ROSEN C A.

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;

DR WPI; 2002-479261/51.
DR N-PSDB; AB084825.

PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
PT and for preventing or attenuating disease caused by Streptococcus
PT infection.

PS Claim 11; Page 29; 70pp; English.

CC AB084792 to AB084904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.

OY 657 YKGELEKGYQPDGMEISGFGKKDAGVYINLSKDTFKKPKFKKEKKEENKPTDVSX 716
DB 1 YKGELEKGYQPDGMEISGFGKKDAGVYINLSKDTFKKPKFKKEKKEENKPTDVSX 60

RESULT 11
ADC45149
ID ADC45149 standard; protein; 117 AA.

XX ADC45149;

DT 18-DEC-2003 (first entry)

XX S. pneumoniae antigenic protein SP043.

XX Antigen; bacterial infection; vaccine; pneumonia; antibacterial.

XX Streptococcus pneumoniae.

XX US6573082-B1.

XX 03-JUN-2003.

XX 28-MAR-2000; 2000US-00536784.

XX 31-OCT-1996; 96US-0029960P.

XX 30-OCT-1997; 97US-00961083.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;

DR WPI; 2003-764574/72.
DR N-PSDB; ADC45148.

PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
PT useful for producing vaccines for prevention or attenuation of infection
PT by Streptococcus pneumoniae.

PS Example 1; SEQ ID NO 68; 58pp; English.

XX The invention relates to an isolated polynucleotide consisting of a
XX Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX antigens. Also included are making a recombinant vector by inserting the
XX nucleic acid into a vector, an isolated polynucleotide consisting of at
XX least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX acid are useful as DNA vaccine against Streptococcus pneumoniae
XX infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX antigen nucleic acids are useful as probes for use in diagnostic methods
XX for detecting S. pneumoniae gene expression. The present sequence
XX represents an S. pneumoniae antigenic protein.

SQ Sequence 117 AA;
 Query Match 15.3%; Score 615; DB 7; Length 117;
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 657 YKGELEKGYOPDGEISGFECKDAGYVNLKDTPIKPVKFKI EKKKEENKRTPVSK 716
 DB 1 YKGELEKGYOPDGEISGFECKDAGYVNLKDTPIKPVKFKI EKKKEENKRTPVSK 60
 QY 717 KCONPVNHSQNLNESHKEDLQREHSQKSDSTQVATVLDKNNISSKSTNNPNK 773
 DB 61 KCONPVNHSQNLNESHKEDLQREHSQKSDSTQVATVLDKNNISSKSTNNPNK 117
 RESULT 12
 ABU24813
 ID ABU24813 standard; protein; 1196 AA.
 AC ABU24813;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DB Protein encoded by Prokaryotic essential gene #10340.
 XX
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
 OS Clostridium botulinum.
 XX
 FN WO200277183-A2.
 PD 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002MO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362599P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zykkind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu RH;
 DR WPI: 2003-029926/02.
 DR N-ESDB; ACA28683.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 XX Claim 25; SEQ ID NO 52737; 1766bp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1196 AA;
 Query Match 5.6%; Score 227; DB 6; Length 1196;
 Best Local Similarity 20.2%; Pred. No. 0.00021;
 Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;
 QY 1 KLGRIASKEKRNGLNGKRG-----SLKKDTTGVETHHQENESI 39
 DB 123 RLNDIOE-LFMDYIGREGYSIIQCKIEAVLSGKPEERSLLEAGIYFKRKXED 181
 QY 40 KEKSSPTIDRISTIPDEN--KDKKLICKTREVDF--TSSTGRME----- 85
 DB 182 KKLNS--TEQNLIRKDIILATYERMEPLKEBSEKAKFNLSELRKKEVNVWISIDK 239
 QY 86 -EYDYKTDKGNIIAVDDGT-DLEYETEKDEIKSKYGLVLSKDGHPFELIGTSVSK 143
 DB 240 IERDILK-NISSNLSIGENDLNKREKSYKEIISK-----NEKILDLK--NNSR 288
 QY 144 NARVYIGANNYSIIEIKATKYDFHSKTFPDLVANNIDYGLAPAGDMRLFVNDPKKA 203
 DB 289 NKEEYYNKKDKNIDR-----NENALKEKIKLKD-----NIKVKENTLKTN 331
 QY 204 EIKIR-----MPEKIKETSEYPV----- 223
 DB 332 EERLQYLYKEKALEKKKINLKEBESYLSIEIGKEDNANNFNKLEKEKLIKLSBE 391
 QY 224 -----SSYGVYIEIGEDL--SKNPDNLTPWESGKYS-----DSEKQYLLK 265
 DB 392 IELNSTSLRNISVWENHIVLEKNDKNNRNSCSYISSININIKTKEDIEKINIK 451
 QY 266 DNILL-----KRGVALKTTVPKGTMLEGNGVYSKED-----IAKIQ-----AN 307
 DB 452 ENLILLENLKENSRNISLKIISLNKKEK-KLKEKNAAYSRLEANYHMLNLEKHYGVN 510
 QY 308 PNLRALSETTIYADSNVDSRSTOSVLSALDGFNIIHQVTFPKGNKGEAIDDGKL 367
 DB 511 RSVKTLMEVSYGKVDNINIG-----CEVLGDIIIKKKELETAIEALGAI-----SNV 560
 QY 368 VTDSSKLVLGKDDKEYTEGDEKFNVEAIEDSGMLPDPKPVLSMDKVFNP-----SKSN 424
 DB 561 ITEDS-----NKAKILNLYLKK-----SLGRATFLPLTTIGR 534
 QY 425 KLYVRPEYTLAKGSKDKGFWELRVNESVVDNYLYGDLIINDTRDRI--KLVNKG 492
 DB 595 KAKINN-----VIREDF--LGIASDLIDY-----DVRSNIIDVLAGRTLVAKD- 637
 QY 483 DIMDWMKDYKANGFPDKTIDMGVNYLQTVSGLNAKAVGHYGLVNVNVEVNIIDPK 542
 DB 638 --MDSKLKAKGLANSFKLVLEGV-----INIG 665
 QY 543 GNT--SIERYADKSVFENINIKRNNFGDEIQOHIIYNGKEYTSFNDIKOIIDTTLNI 599
 DB 666 GSLTGGSIGIRAGSSI--ISRR-----RIE-----TKBELBETN- 700
 QY 600 KLVVDPAANTYKFEPIINKONGEVSLEKPHRVYTTIQNGKEMSSTIVEEEDPIIPVYG 659
 DB 701 -----TIEEFMEK-----ILEKNKIKT--LDEENLNI--KD 728
 QY 660 BLEKGYOPDGEISGFECKDA-----GVVNLKSD--TFIKPVFKIEKKKEENK 709

PR 27-AUG-2002; 2002US-0406676P.
 PR 28-AUG-2002; 2002US-0406757P.
 XX
 PA (CHTR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Tettein H, Masjmani V;
 XX
 DR WPI, 2004-248071/23.
 XX
 PT Immunogenic composition useful as a vaccine for treating or preventing
 PT streptococcal infections, comprises group B Streptococcus polypeptides.
 XX
 PS Claim 10; SEQ ID NO 4422; 1194pp; English.
 XX
 CC The invention relates to a novel immunogenic composition comprising a
 CC combination of 2-5 group B Streptococcus (GBS) polypeptides. Each
 CC polypeptide is encoded by a GBS polynucleotide sequence which is
 CC homologous to a polynucleotide sequence of group A Streptococcus (GAS),
 CC Streptococcus pneumoniae and/or at least one other GBS serotype. The
 CC composition of the invention demonstrates antibacterial activity whilst
 CC the polypeptides and polynucleotides may be useful in assays to diagnose
 CC and identify streptococcal infections or for identifying, screening and
 CC developing vaccines and other treatments for streptococcal infections.
 CC The current sequence is that of a Streptococcus agalactiae ORF SAG
 CC protein of the invention.
 CC
 SQ Sequence 1166 AA;

Query Match 5.5%; Score 220; DB 8; Length 1166;
 Best Local Similarity 20.2%; Pred. No. 0.00051;
 Matches 198; Conservative 139; Mismatches 306; Indels 336; Gaps 56;

QY 4 ELAEEKPKNLKGKSGSLKQTTGVHHQENBESIKKSSPTIDNISTINDFENKOLK 63
 DB 60 ELLEELSKNLDTSNMCA-----DLSEKPSKPEITNNKES-NVNTNASTIA-----IA 105
 QY 64 KLIEKKKPREVDPTF-----ETGKMEBYDYKDDKGNIA-YDDGDLLEYETKLDDEI 116
 DB 106 QKVPAYBEVKPKSSSLAVLDTSKITKQAITQKRGKVVAVLIDGPFINDIFLDD-- 163
 QY 117 KSKITGVLSPKDGH-----PELLGKISNVS---KNAKYTYGNVYS---LEIYA 160
 DB 164 -----SPKDDKHSFKKAKAFEBELKAKNITYGKVVNDKIVFAHNVANNETVADIA 215
 QY 161 TKYDHSKMTFDTLVANINDIVDGLAFAGMELFVNDONOKA-----EIK 206
 DB 216 AMKDGIG-----SEANNISHGTHVAG--TFV-GNSKRPAINGLLEGAAPNAQVL 263
 QY 207 IRMPKIKETKSEYFVSSYGVNVIELGSDLSKNKPDNLTKNESGR-----IYDSERKOQY 262
 DB 264 MRIPKIDSDKFGBAVAKAITDAVNLGATIN-----MSIGTASLIALANDKVL 314
 QY 263 LKKNIIILKGYALVYTTNPGKTMLBGNVYSKEDIKIQGANP-----NLRAISPTT 317
 DB 315 ALK--LASERKGYAVVVAAGNBAFGM-----DYSEKPLSTNDPYGVNSPAISEDT 362
 QY 318 IYADSRNVEDGRSTQVLSALDGFNIIRYOVFTPGONKRGABIDKDGMLVYDSSGLVVF 377
 DB 363 LSVAS--YESLKTISEVATTIEG-KLVKLPVITSKPFNGKAYD-----VVY 407
 QY 378 GK--DDKRYTGD--KENVAIKEDGMLFID--TKFVNLSMDKRYFNPSKNTIYRANB 432
 DB 408 ANYGAKDFEGDFKQKALIERGGGLDMTKITHATNGV-----VGIVAFENDQ 457
 QY 433 FYLRKIDSDKGFPMNELRVNESVDNYLI-YGDL-----HIDNTRD--FNITL 477
 DB 458 -----EKRG-----NPLPYRELPGVATSKYDGERIKNTISQGLTGNQRP 496
 QY 478 NVKD--GDIM-----DMGM-----KDYKANGFPDRTMDQGNVLIQTGYSDLNKAVG 523
 DB 497 EYVDSGGNRMLEQSSWSVTAAGAIKPDVTASGFEIYSSTYNNQYPTMGSTMASHPVAG 556

QY 524 V-----HYQFLYDNYKPEVNTIDPK-----GNTSIEYADGKSVFVNIDKRNNG 566
 DB 557 LMTMLQSHLAERKKG-----MNLDSKLLBELSKNLIWSSATPALYSBEDKAFYSRQGGAGV 612
 QY 567 PDGE--IQDQHIYINKERTSFNDIKQIITDKTANIKIYVDPFARNTTKEFPIINKDTGEV 624
 DB 613 VDAEKAIQAOY-VYVNDGKAKINLRVGDK-FDITVTIHKLVGCVKELYQANVATEQV 670
 QY 625 SE--LKPH-----RYVTITQ-----NG----- 639
 DB 671 NKGKPAKFAQALDNTWQVILRDKETQVRFITIDASQFSQKLKQVANKYFLEGVPRPK 730
 QY 640 -XEMSSIVS-----BEDFI-----LPYKGELEKGY-----QPDGEISG 674
 DB 731 AKDSNQLMSTIPVGRNPFASIQALEPIYK-TLSKSPYYPKPNTHDQLEYNESAP 789
 QY 675 FEGSK-----DAGYV-----INLSKT--FKPVPK-KIEEK-----KEE 707
 DB 790 FESNNYTLALTQSASGWSYVDYKNGGELIAPESPRIILIGPENKVEDTTHILRDA 849
 QY 708 NKPTFVSKKXDPQVNSQLNBSHXKEDLOREHSQXSD--STRDVTATVLDKRN-NI- 762
 DB 850 NMPYFALSFKDG-----NRDEITPQATFLANVKDISAQVLDQNGVNI 892
 QY 763 -----SKSTNNPNK 773
 DB 893 WQSKVLPSYRKPNFNNPKQ 911

RESULT 15
 AAM54145
 ID AAM54145 standard; protein; 1639 AA.
 XX
 AC AAM54145;
 XX
 DT 25-MAR-2003 (revised)
 DT 23-SEP-1998 (first entry)
 XX
 DE P. falciparum synthetic gp190 protein.
 XX
 KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
 KW monoclonal antibody; passive immunisation; parasite.
 XX
 OS Plasmodium falciparum.
 OS Synthetic.
 XX
 PN W09814583-A2.
 XX
 PD 09-APR-1998.
 XX
 PF 02-OCT-1997; 97MO-BP005441.
 PF 02-OCT-1996; 96DE-01040817.
 PR (BUQA/) BUJARD H.
 PA Bujard H, Tolle R, Pan W;
 PI WPI, 1998-240088/21.
 DR N-PSDB; AAV21451, AAV35363.
 DR
 XX
 PT Recombinant production of complete gp190/MSP-1 Plasmodium surface protein
 PT - useful in anti-malaria vaccines, also stabilising genes by reducing
 PT their AT content.
 XX
 PS Example 1; Fig 3c; 48pp; German.
 CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
 CC (merozoite surface) protein. The gene encoding this protein has been
 CC stabilised by reducing the AT content of the nucleotide sequence. Such a
 CC protein is useful in vaccines against malaria or for producing monoclonal
 CC antibodies (for passive immunisation). The complete gp190 protein can now
 CC be produced outside the parasite and has, at least over extended regions,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:10:40 : Search time 45 Seconds
(without alignments)
1282.305 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGIASKEKFNILGNKSGS.....ATVLDKNNISKSSTNNPNK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3789.5	94.1	2138	4	US-09-583-110-5274
2	3048	75.7	637	4	US-09-107-433-3169
3	750.5	18.6	1529	4	US-09-107-433-4771
4	615	15.3	117	3	US-08-961-083-68
5	615	15.3	117	4	US-09-536-784-68
6	203	5.0	1166	4	US-09-200-650E-7
7	200.5	5.0	1786	3	US-08-973-462-8
8	191.5	4.8	2504	4	US-09-328-352-5821
9	188.5	4.7	670	4	US-09-107-433-4976
10	188.5	4.7	708	4	US-09-583-110-1019
11	187.5	4.7	1529	2	US-08-728-470-10
12	187.5	4.7	1529	3	US-08-719-641-10
13	186.5	4.6	930	4	US-09-200-650E-3
14	186	4.6	1315	4	US-09-200-650E-5
15	185	4.6	2733	4	US-09-949-016-11433
16	185	4.6	3259	4	US-09-949-016-6507
17	184	4.6	1588	5	PCT-US93-07261-11
18	184	4.6	1663	5	PCT-US93-07261-16
19	181.5	4.5	778	3	US-09-134-001C-3866
20	179.5	4.5	746	4	US-09-710-279-652
21	179.5	4.5	1183	2	US-08-447-031A-2
22	179	4.4	2142	4	US-09-540-236-3459
23	178	4.4	2375	4	US-09-538-092-1131
24	177.5	4.4	1600	2	US-08-617-697-10
25	177	4.4	800	6	5183745-3
26	177	4.4	800	6	5183745-3
27	174.5	4.3	1073	4	US-09-206-942-49

28	174.5	4.3	1079	4	US-09-206-942-47	Sequence 47, Appl
29	174.5	4.3	10182	3	US-09-134-001C-3159	Sequence 3159, Ap
30	174	4.3	3070	4	US-09-961-403-7	Sequence 7, Appl1
31	173.5	4.3	2353	3	US-09-377-155-33	Sequence 33, Appl
32	173.5	4.3	2353	3	US-08-913-942-4	Sequence 4, Appl1
33	173.5	4.3	2353	3	US-08-669-974-33	Sequence 33, Appl
34	173.5	4.3	2353	4	US-09-797-862-33	Sequence 33, Appl
35	173.5	4.3	2353	4	US-09-684-707-4	Sequence 4, Appl1
36	173.5	4.3	2411	3	US-09-268-347-36	Sequence 36, Appl
37	173	4.3	930	4	US-09-134-001C-5314	Sequence 5314, Ap
38	173	4.3	930	4	US-09-386-962C-10	Sequence 10, Appl
39	173	4.3	2710	1	US-08-480-604A-6	Sequence 6, Appl1
40	173	4.3	2710	2	US-08-405-496A-6	Sequence 6, Appl1
41	173	4.3	2710	3	US-08-915-136-6	Sequence 6, Appl1
42	173	4.3	2710	4	US-08-957-310-6	Sequence 6, Appl1
43	173	4.3	2710	3	US-10-011-366-6	Sequence 6, Appl1
44	173	4.3	2710	4	US-09-084-517-6	Sequence 6, Appl1
45	172.5	4.3	1092	4	US-09-147-405B-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-583-110-5274

Sequence 5274, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al.

TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus

TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics

FILE REFERENCE: PATHO0-07A

CURRENT FILING DATE: 2000-05-26

PRIOR FILING DATE: 1998-06-30

PRIOR FILING DATE: 1998-05-12

PRIOR APPLICATION NUMBER: US 60/085,131

PRIOR FILING DATE: 1997-07-02

NUMBER OF SEQ ID NOS: 5322

SEQ ID NO 5274

LENGTH: 2138

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-583-110-5274

Query Match	94.1% Score 3789.5; DB 4; Length 2138; Best Local Similarity 94.8%; Pred. No. 1.1e-238; Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;
QY	1 KLGIASKEKFNILGNKSGSLKOTGVYHHQNEBSIKESSTTIDRNTTIDPFNK 60
DB	1333 KLGIASKEKFNILGNKSGSLKOTGVYHHQNEBSIKESSTTIDRNTTIDPFNK 1392
QY	61 DLKLLIKKKFPRVDDFTSRTGKMEBYDKYDDKGNIIAYDDGTLEYTEKLDKISKI 120
DB	1333 DLKLLIKKKFPRVDDFTSRTGKMEBYDKYDDKGNIIAYDDGTLEYTEKLDKISKI 1451
QY	121 YGVLSPSRSDGFELIGKISVSKAKVYGGNNYSIRIKATKYDFHSKMTFDLYANIND 180
DB	1452 YGVLSPSRSDGFELIGKISVSKAKVYGGNNYSIRIKATKYDFHSKMTFDLYANIND 1511
QY	181 IVDGLAFPGDMRLFPKNDOKKAKIKIMPEKIKETSEYVSVYGVNIEGEGDLSKN 240
DB	1512 IVDGLAFPGDMRLFPKNDOKKAKIKIMPEKIKETSEYVSVYGVNIEGEGDLSKN 1571
QY	241 KPDNLTAKESGKISDSERKOOYLAKDNILIRKGYALKTTYNPGKTDMLBENGYSKEDI 300
DB	1572 KPDNLTAKESGKISDSERKOOYLAKDNILIRKGYALKTTYNPGKTDMLBENGYSKEDI 1631
QY	301 AKIQKAPNPLALSETTYIADSRNVEDGRSTQSVLMSLDGFNITRIYQVTFKQNDKGEA 360

```

Db      1632 AKIQKANPNLRALSETTITYADSRVVEDGRSTQAVLMSALDGFNIIIRYQVFTFRANDGBA 1691
Qy      361 IDKGNLVTDSSKLVLFQKDKKEYTSGDKFNEVAIKEDGSMFLPDTKPVNLSDMKXNFPN 420
Db      1692 IDKGNLVTDSSKLVLFQKDKKEYTSGDKFNEVAIKEDGSMFLPDTKPVNLSDMKXNFPN 1751
Qy      421 SKSNKIYVRPPEFLRKISIDKGGFNMELRVNESVUNYLLYGLHIDNTRDPNFKLVNK 480
Db      1752 SKSNKIYVRPPEFLRKISIDKGGFNMELRVNESVUNYLLYGLHIDNTRDPNFKLVNK 1811
Qy      481 DGDIMWGMDKYAMGFDPKVTMDGNNVYLQTYGSDLNAAKGVHGYOFLYDNVPEVNID 540
Db      1812 DGDIMWGMDKYAMGFDPKVTMDGNNVYLQTYGSDLNAAKGVHGYOFLYDNVPEVNID 1871
Qy      541 PKGRTSIEYADGKSVFNINDKRNNGPDGEIOEQHIYINGKEYTSFNDIKOITDKTLNFK 600
Db      1872 PKGRTSIEYADGKSVFNINDKRNNGPDGEIOEQHIYINGKEYTSFNDIKOITDKTLNFK 1931
Qy      601 IVVNDPARNNTVKEFLNKDTGSEVSELKPHRVTVTTIONGKEMSTIYSEEDFILPVYKGE 660
Db      1932 IVVNDPARNNTVKEFLNKDTGSEVSELKPHRVTVTTIONGKEMSTIYSEEDFILPVYKGE 1991
Qy      661 LEKGQFDGWEISGFEGKQAGYVNLKDTPIKPVFKKIEKKKEENKPTFDVSKKKN 720
Db      1992 LEKGQFDGWEISGFEGKQAGYVNLKDTPIKPVFKKIEKKKEENKPTFDVSKKKN 2051
Qy      721 POWNHSLMNSHRKEDLOREHSHQKSDSTQVTAIVLDKNNISSKSTNNPNK 773
Db      2052 POWNHSLMNSHRKEDLOREHSHQKSDSTQVTAIVLDKNNISSKSTNNPNK 2104

RESULT 2
US-09-107-433-3169
; Sequence 3169, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOMER THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULAR TYPE: Protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: m1ec feature
; LOCATION: (8) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-433-3169

Query Match      75.7%; Score 3048; DB 4; Length 637;
Best Local Similarity 99.3%; Pred. No. 5,6e-191;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy      185 LAFAGDMRLPVKNDQKASIKIMPEKIKETKSEYTVSSYGVNVELEGSDLSKRYKDN 244
Db      15 LAFAGDMRLPVKNDQKASIKIMPEKIKETKSEYTVSSYGVNVELEGSDLSKRYKDN 74
Qy      245 LTKHSGKITYSDSKQOYLKNDIILRKGYALKYTYNPGKTDMLGNGVYSKEDIKIQ 304
Db      75 LTKHSGKITYSDSKQOYLKNDIILRKGYALKYTYNPGKTDMLGNGVYSKEDIKIQ 134
Qy      305 KANPNLRALSETTITYADSRVVEDGRSTQAVLMSALDGFNIIIRYQVFTFRANDGBAIDKD 364
Db      135 KANPNLRALSETTITYADSRVVEDGRSTQAVLMSALDGFNIIIRYQVFTFRANDGBAIDKD 194
Qy      365 GNLVTDSSKLVLFQKDKKEYTSGDKFNEVAIKEDGSMFLPDTKPVNLSDMKXNFPNPSKN 424
Db      195 GNLVTDSSKLVLFQKDKKEYTSGDKFNEVAIKEDGSMFLPDTKPVNLSDMKXNFPNPSKN 254
Qy      425 KLYVRNPEFLRKISIDKGGFNMELRVNESVUNYLLYGLHIDNTRDPNFKLVNKDGI 484
Db      255 KLYVRNPEFLRKISIDKGGFNMELRVNESVUNYLLYGLHIDNTRDPNFKLVNKDGI 314
Qy      485 MDKGMKQYKANGFPDKYTTMDGNNVYLQTYGSDLNAAKGVHGYOFLYDNVPEVNIDPKN 544
Db      315 MDKGMKQYKANGFPDKYTTMDGNNVYLQTYGSDLNAAKGVHGYOFLYDNVPEVNIDPKN 374
Qy      545 TSIEYADGKSVFNINDKRNNGPDGEIOEQHIYINGKEYTSFNDIKOITDKTLNFKIYVK 604
Db      375 TSIEYADGKSVFNINDKRNNGPDGEIOEQHIYINGKEYTSFNDIKOITDKTLNFKIYVK 434
Qy      605 DPARNNTVKEFLNKDTGSEVSELKPHRVTVTTIONGKEMSTIYSEEDFILPVYKGELEKG 664
Db      435 DPARNNTVKEFLNKDTGSEVSELKPHRVTVTTIONGKEMSTIYSEEDFILPVYKGELEKG 494
Qy      665 YQFDGWEISGFEGKQAGYVNLKDTPIKPVFKKIEKKKEENKPTFDVSKKKNPQVN 724
Db      495 YQFDGWEISGFEGKQAGYVNLKDTPIKPVFKKIEKKKEENKPTFDVSKKKNPQVN 554
Qy      725 HSQNLMSHRKEDLOREHSHQKSDSTQVTAIVLDKNNISSKSTNNPNK 773
Db      555 HSQNLMSHRKEDLOREHSHQKSDSTQVTAIVLDKNNISSKSTNNPNK 603

RESULT 3
US-09-107-433-4771
; Sequence 4771, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOMER THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4771:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1...1529
SEQUENCE DESCRIPTION: SEQ ID NO: 4771:
US-09-107-433-4771

Query Match 18.6%; Score 750.5; DB 4; Length 1529;
Best Local Similarity 79.7%; Pred. No. 1.8e-40;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

QY 1 KLGEIASEKFKLNGKSGSLKKDTGVGHHQENESIKESFTIDNISTIDPEK 60
DB 1336 KLGEIASEKFKLNGKSGSLKKDTGVGHHQENESIKESFTIDNISTIDPEK 1395
QY 61 DLKLLIKKKFVDDFTSEYTGKMEYDYKYDDKNIIAYDDGTDLEETEKLDISKKI 120
DB 1396 DLKLLIKKKFVDDFTSEYTGKMEYDYKYDDKNIIAYDDGTDLEETEKLDISKKI 1454
QY 121 YGVSPSGDGHFRILGKISNYSKNAKYVYNNYKSIETKATYDDHSHKMTEDLYANIND 180
DB 1455 YGVSPSGDGHFRILGKISNYSKNAKYVYNNYKSIETKATYDDHSHKMTEDLYANIND 1514
QY 181 IVDGLAF 187
DB 1515 IVDGISF 1521

RESULT 4
US-08-961-083-68
Sequence 68, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-68

Query Match 15.3%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.9e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YVGELEKGYQPDGWEISGFGKDDAGYVNLNKKDFIKVFVKLEKGBENKPTDVS 716
DB 1 YVGELEKGYQPDGWEISGFGKDDAGYVNLNKKDFIKVFVKLEKGBENKPTDVS 60
QY 717 KDNFQVNSQLNESHKEDLQREHSQKSDSTQVATVLDKNISKSTNNPK 773
DB 61 KDNFQVNSQLNESHKEDLQREHSQKSDSTQVATVLDKNISKSTNNPK 117

RESULT 5
US-09-536-784-68
Sequence 68, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:

```
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match      15.3%; Score 615; DB 4; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.9e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGLKGYQDPGRIGSGFGKQAGYVYINLSKQFTICVPRKIEKKEEENKPTPDVSK 716
DB 1 YKGLKGYQDPGRIGSGFGKQAGYVYINLSKQFTICVPRKIEKKEEENKPTPDVSK 60

QY 717 KQNPQVNHQSQANSRHRKEDLQREHSQKSDSTKQVATVLDKNNISKSTNNPNK 773
DB 61 KQNPQVNHQSQANSRHRKEDLQREHSQKSDSTKQVATVLDKNNISKSTNNPNK 117

RESULT 6
US-09-200-6508-7
Sequence 7, Application US/092006508
Patent No. 6680195
GENERAL INFORMATION:
APPLICANT: Patti, Joseph M.
APPLICANT: Foster, Timothy J.
APPLICANT: Hook, Magnus A.O.
APPLICANT: Eichlim, Delidre M.
APPLICANT: Perkins, Samuel L.
TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
FILE REFERENCE: P06283US2/BAS
CURRENT APPLICATION NUMBER: US/09/200,650E
PRIOR FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066,815
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: 60/098,427
PRIOR FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1166
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-200-6508-7

Query Match      5.0%; Score 203; DB 4; Length 1166;
Best Local Similarity 21.6%; Pred. No. 6.8e-05;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

QY 33 QENESLIKE-KSEFTIDRNISTIRDFENKDLKIKKKR-----EYDDPTS 78
DB 226 KNNBKKKELVRNDSMTDSTKPAVTAAPKVAAPKVAQAQAASNNVNDLIK 285
QY 79 EYGRMEBYDYKDDKGNIIAYDDGTDLEETE-----KDEIKSKIKYGLS 125
DB 286 VTKGTIK--VGGCKDNVAALHDKDIEYDEFTIDKKYKKGDMTINIKNVIPSDLT 341
QY 126 PSKC-----GHPEILGISNVSKNK---VYGNKYKSIKATYDFHSK----- 168
DB 342 DKNDPIDITDPSGEVIAKGTEDKATKQITFTVDVVDKEDIKSLTLTYSYIDKKTVPMB 401
QY 169 ---TMTF-----DLVANTN-DIVDELAFAGD---MRLVVK-DNDQKKEIKIRM-PEKIK 214
DB 402 TSLMLTFRATKAKESQNTVYQDPVNH-GDSNIGSITFKDDEBKQTEQOIYVNPPLKS 460
QY 215 EYKSEY-----PYVSSYGNVIELGSGD--LSKQKPDNLTYM-----ESGKIYSDSKQ 261
DB 461 ATNTKVDIAGSQVDYGN-ITLNGSSTIIDQNTBEIKYKVNSDQQLPSNNIYDSQYED 519
QY 262 YLLK-DNIIILKGYALKATYTNPKTIDMLGNGVYSKEDIKIQANPNPALSETTIIYA 320
```

```
DB 520 VTSQFDN---KKSFSNNVALIDFG-----DINSAYI-----IKVSKTTPS 558
QY 321 DSR-NVEDGRSTQVLSALDGFNIRQVFTFRKNDK-----E 359
DB 559 DGBELDIAGTSMRTT--DKYGYVYVAGYSNFIYVSNIDYGGGQGVKPRKLYKIDGVYWE 616
QY 360 AIDKQNLVYDSSSLVLFKQDKKEYGDKXNVALKRDKGSMFLPDTQPVNLSMKNY-F 418
DB 617 DVDKDG-----VQSTSKKCPMANVL-VLTYPDGT-----TYSVRTDAGHYEF 660
QY 419 NPSKSNKIYV---RNPEFYLRKIS-----DKGCFMWEJRNVESVVDNYLIYDGLHD 468
DB 661 GGLADGETTYVKEPIPTCYLPTKNGITDGERKDSGSSVYTKLNK-----DDMSLD 712
QY 469 -----NTRDFNLIKLVNDG--DINDWGRQYKANGPPDKVTDMQGNVYLQGYSDL 517
DB 713 TGFYKEPKYVNLGDYWEEDYKQIDANBPGIKDKVLT-----LKDSYGVN-IGTTTIDA 766
QY 518 NAKAVGNYQPLVNDVKEPVNIDPKGNISIEY---ADGKSVFN--INDKRNNGFDGELQ 572
DB 767 SGK-----YKF-----TDLD-NGNTVVEPETPAGTTPVKNTTADDKOSNGL----- 807
QY 573 EQLIYINKEEYTSFNDIKQIIDKTLNIRKIVKDFAPANT--TWKEPI---LNKDTGVBSEL 627
DB 808 -----TTTGVIDADANMTLD-----RGFYKTPKYSGLGDVYVYDSDNKGKQDSTE 851
QY 628 KPRH-VYTTIONGR-ENSSITVSEEDFLPYKGLKGYQDPGRIGSGFGKQAGYVI 685
DB 852 KQIKDVTYVTLQNEKEEVIQIKTDEN-----GR-----YRFD----- 883
QY 686 NLSKDTPIKVPKIKKEKEEENKPTPDVSKKQNPQV-----NHSQ--LNESHKEDLQ 738
DB 884 NLDGSKT-KVIFRKPAGTQVYVTTTBD-DKADAGGADVYITTDHDTFLDNGYFEEDTS 941
QY 739 REEHSQKSDSTKD 751
DB 942 DSDSDSDSDSDSD 954

RESULT 7
US-08-973-462-8
Sequence 8, Application US/089734628
Patent No. 6191270
GENERAL INFORMATION:
APPLICANT: DRUTHE, PIERRE
APPLICANT: DAUBERSIES, PIERRE
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,462B
PRIOR FILING DATE: 1998-02-06
PRIOR APPLICATION NUMBER: PCT/FR96/00894
PRIOR FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007
EARLIER FILING DATE: 1995-06-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1786
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match      5.0%; Score 200.5; DB 3; Length 1786;
Best Local Similarity 20.5%; Pred. No. 0.00018;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

QY 6 AAS--KFNK-LANGKESLKKQD-----TGVVEHQENR-----ESIK 40
DB 868 AESTVTPNIIIEIQTNTTDTIRKLEBLHENVLSAALNTQSEEEKGVIVDIYIRVK 927
```



```

QY 41 EKSPFI-----DRTSTIRD-PEN-----KOLKLIKKKPREVDFTS 78
DB 928 BEVATTLITVEOAEKSKSANTLTETFEENAEVSENVAEKLKLNRTVFTVLDKVE 997
QY 79 ET-----GKAME--EYDYKY-----DDKG--NIIAYDSTDLLEY-----EKLD-- 114
DB 988 ETVEISGBELNNENDKAFPSIFDVKIGQENLTL--GMPFSIFTSIVIOSEKVDLN 1044
QY 115 -EIKSKIYGVLSPSKDGHEILIKISNVSRNAVY-----YGNKYKSIETATYDFH 166
DB 1045 ENNVASSILNINIKMKKG---LANKLENISTSGVETVEHQVNYVVDVDAK--- 1097
QY 167 SKTMTFPLVANIINDIVIDGLAFAGMRLPYKNDQKKAELIKIRPEKIKETKESPYVSSV 226
DB 1098 -----DQFLGILNAGL-----KEMFPLEDPVPSBSVDIVTEBEIKOEPVKEVEKET 1146
QY 227 GNVIELGSD-----LSKNRPMULTRM-----ESGKIYDSEKQOYLKONILIRKGYALK 277
DB 1147 VSIIEMENIIVDVEEKEBDLTDKMLDAVEBSIEISSQKETSIGKOK--EKQVSLV 1203
QY 278 VTTNTPG-----KTMMLBGNVYSK--EDIAKIQKANPUL----- 310
DB 1204 YEEVODNDDESVEKVLBNNEBELMKDAVEINDITSKLIETOBLINEVADLIDOMEK 1263
QY 311 -----RALSETTIYADSRNVEDGR--STOSVLSALDGFNIIIRYQVFTFQNDKGEALDK 363
DB 1266 LKELEKALSE-----DSKEIIDAKDPTLEKVIIEED-----ITTTLDVVELKDV 1309
QY 364 DGNLVTSSKLVLPFGDKDEYTGEDKFNVEBAIKEDSGMLFIDTKPVNL-----SM 413
DB 1310 EEDKLEKVDL-----KOLBEDILK--VKEIKELSEILEDYKBLIETDILBEKKEI 1362
QY 414 DKNYRNPSSNCTIYANRPFYLRGKISDGKGFMEALVNESVVD-----NYLYGDL 465
DB 1363 EKHHEKFEKEEBAEIKDLADILKEVS-----SLVEBEKQLEVEHLEKEVEHIIISGA 1417
QY 466 HINDTRDFNIKLVND--GDIM-----DWGKQDYKANGPPDKVTMDGNYVLTQYSDLA 518
DB 1418 HIKGLBEEDLE--EVDILKGSILDMKGMELGMDGSELEDYTKGGERV--BSLKQVL 1473
QY 519 AAVGVHYOPLYND-----VKPEVNIDPRGNTSIRVADKSVFVNINDK--RN 564
DB 1474 SSALGDEBQOMTKRKAQRKLEEVLLKEEVEKEPEKKIT-----KKQVRFIDKDEPKD 1528
QY 565 NGFDEIIOBHITYNCK-----EYTSNDIKQIIDKTLN-----IKLVNDPFR--N 609
DB 1529 BIVEVEMKOEDIEBDEVEDIEEDIEEDKVEDIDEDIDEGEDKVEDIDLIVKEKRIK 1588
QY 610 TTVKEFIILNKDGE--VSELKPH-----RVTYTIQNGKEMSSTIVSEED 651
DB 1589 VAAKKKKLEKCYBEGVSGIKGVDEVMKVQKIDKEVDKEVSKALESKNDVTNVAKQND 1648
QY 652 PLYPYKGLKGYOPDGEISGFBGSKDAGYVINI-----SKDTFIKPVK 698
DB 1649 PFSKY-KNVVKKYKFAAPFISAVAA--PASYVVGFTFSLFSSCVTIASSYTLISKVD 1705
QY 699 KLEBKKEENKX-----TPDV-----SKKDNQPVNSQUNESRKEEDLQR 739
DB 1706 TINKNE--RPFYSVFPIFPLNKLHYLOOMKEKESKEKNNTYEVTL--NKEKKGNVQV 1760
QY 740 EHSQKSDSTKDVATVLDKNKNISSKSTNNPK 773
DB 1761 TMTKTEK-----TKVDKNNKVPKGRRTQSKS 1786

```

RESULT 8

```

US-09-328-352-5821
; Sequence 5821, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

```

```

; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5821
; LENGTH: 2504
; TYPE: PRN
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-5821

Query Match      4.8%; Score 191.5; DB 4; Length 2504;
Best Local Similarity 19.5%; Pred. No. 0.0011;
Matches 198; Conservative 130; Mismatches 346; Indels 341; Gaps 54;

QY 5  IAEKRNKANGKESLKQDTGVENHHQENESIKESFTID-----RNISTIDPE 58
DB 1365 LKDSKEDSV--TYGNTTLATNGL-----TIKGPSTYEGINAGSKQITNVADGI 1412
QY 59  NKOLKLIKPPREVDFTSENGKMEBYDYKDD-----KGNIIAYDSTDLLE-- 107
DB 1413 N-AKDAVNS--QDNLAKQ--NATDDAAVKYTDKATKQKVLTKK--DGVLDNV 1462
QY 108  ---YETEK-----LDEIKSKIYG--VLSP-----SKDGHEILIKX 137
DB 1463 KAGHISTSKBAVNSQIHKISINSIGANTVAVPDGSLTNNIGTGKNNINDALISE 1522
QY 138  ISNVSKAK--VYGNNTYSIEIK-----AKTDFHSK-----TMTFPLVANIINDI 181
DB 1523 VKNATKATKATYTBGN--IYKETEYVNDGSTRYEVATKQDVLNSVTTGDTVLANNGL 1579
QY 182  V-----DGLAFAGMRLPYKNDQKKAELIKIRPEKIKETKESPYVSSYGVNT-- 230
DB 1580 TTYDGSITKDGINSKQITNVADGINAKDAVNVQDLTKYKELN--GRITDTNNGLND 1637
QY 231  ---ELGEGDLKSNKPNMLTKMESGKIYSDSEKQOYLKONILIRKGYALKVTTYNPKTD 287
DB 1638 AKKQDLNQJADTKNLNDAKQDLGNQITDTNTLTANTTKQDL-----TYQINDKTE 1688
QY 288  MLEBNG-----VSKEDIAKIQKANPRLR-----ALSETTIYA 320
DB 1689 LNNITGNITELNSKIDSTTELENGKLNPAAGSGNDVHRKLGKRLNIIGGAASTVAK 1748
QY 321  DS-----RNVEDGRSTOSVLSALD-----GFNIIRYQVFTFQNDKGEALIDDG-- 365
DB 1749 TSGENVYTTTQDGIQELKSKSPSVTTGNTTLATNGLTIK--EGPSVTGEGINAGG 1805
QY 366  -----MLVYDTS-----KLVLPKGD----- 380
DB 1806 KKITNVADGVNAKDAVNSQLDNLAAKQNAATDDAAVYDDAKTKDKVTLKQKQGVLDNV 1865
QY 381  ---DKB--YTGEDKRVN-----ALDEGSMLEFIDTKPVNLSMDKXVFN-- 419
DB 1866 KAGHISTSKBAVNSQIHKISINSIGANTVAVPDGSL--LTTNNIG--GTGKNNIND 1921
QY 420  -----PSKSKYIYVRNPEFYLRGKISDGKGFMEALVNESVVDVYLLYGDHLINDTR 471
DB 1922 AISEVQATKATKATYTBEGNIVYKGTAVKDGSTNVEVSTKQDLTNSVTTGDTVLANN-- 1980
QY 472  DFNILKLVNDGDIIMGMDYKANGPPDKVTMDGNAVYLTQYSDLAHAKAVGVHYOPLYD 531
DB 1981 ---GLTIKDGPSI--TKD--GVNAGGKLTVDANKVITAQNSKDAVNA--GAOVHHISN 2028
QY 532  NYKPEVNIDPRGNTSIEVADGKSVFNINDKRNNGPFGIIOBHIIYNGKEVTSFNDIKQ 591
DB 2029 STNNSIG--GNTVNA--PDGSLTNNIGTGKNNINDAKS--VDEKVTNGVNDLTQ 2079
QY 592  IIDKTLNIKLVNDPARN--TYKEPILNKDGEVSELKPHRYVTYTIQNG-----KS 641
DB 2080 ---KGLN-----FANDQKTTQGAVVRKGLD-----TINIVGADAKTAEKDT 2120
QY 642  MSSTIYSB--EDPILVYKGLKGYOPDGMELISG-----EGK 678
DB 2121 SGENIITRTTEBDV--KIEMLKDVKFSVNVGAVLANQGLIKGSPSITVNGINAGG 2176

```

QY 679 KDAGV---INISKTFIK-PVFKIIBENKEENKPTVDVK-----KDNPOV 723
DB 2177 KQITVADGIN-AKDAVNGQIDKQINEVKDIGK---DIGLSHNAVQYDKDKNGVDK 2232
QY 724 NNSOLNESHREBDQREHSHQSDSTKQV-----ATVLDKNNISKSTNN 770
DB 2233 NSVTLGGEGKTNLKNVADGKRVASKQAVNGQIMNIQNVQDKNSNDIKNIQNN 2287

RESULT 9
US-09-107-433-4976
Sequence 4976, Application US/09107433
Patent No. 6800744
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESSES:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Maitland
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ainiello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4976:
SEQUENCE CHARACTERISTICS:
LENGTH: 670 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...670
SEQUENCE DESCRIPTION: SEQ ID NO: 4976:
US-09-107-433-4976

Query Match 4.7%; Score 188.5; DB 4; Length 670;
Best Local Similarity 21.4%; Pred. No. 0.00029;
Matches 138; Conservative 92; Mismatches 233; Indels 183; Gaps 33;

QY 223 VSSYGVNVELEGBDLSKRPDLTMSGKLYSDSEKQY---LKDNIILR--KGYALK 277
DB 54 ISSNGTI-----RSNQDLNRT-VES-TVSTVENKSYKEDVLSDRILKKEPFDALS 104
QY 278 VTTN-----PKTDLBNGNGVYSKEDIKIQ-----KANPNLRLSE 315
DB 105 VKDIGAVGDIHDDRQALIDAAAGGLGGGAVTFPGGTIVLKEIVFLKSHTHLELNEK 164

QY 316 TTVVDSRNVEDGRSTQSVLMSALDGFNIIRYOVFTFRNDKGBAIDKQNTVTDSSKLY 375
DB 165 AT1-----LGINIKNPSIVFWT-----GLPTDGAQV 193
QY 376 LFG-KDDKEYNGEDKFNVEAIKEDGSMFLIDKPNVLSM--DKYFNPSKNTKYRNP 432
DB 194 EMEPTEDISTSGGTDIMNGALNBBG-----TKAKPLFINSNGAPALGNSNNVTIKNT 247
QY 433 F---YLKGIKSDKGFFNMELRVNESVVDNTLYGDLHIDNTEDFNFKLVKRGDIN---D 486
DB 248 FQDSYQGAHIGLAGSKN-----VLVNSRRFLGA-----LPKTMQOQIISK 291
QY 487 WEMKDYKANGFPDXTDWDG---NVTLOTGY---SDLNAK---AVGATQPLYDNVKE 536
DB 292 IOIEPLTRGFPFYALND--DGKSEVNTYONSFYGSDSGSELVTALGTHYQTLSTQPN 350
QY 537 VNIDPKGNTSIEYADGKSVFVNINDKRNNGFPGBIOEHY-----INGKETSFND 588
DB 351 IKILNNHPDNNMTAGVRFPGFTDVLKGRFPAKKGESVHRESGALVNAYSTNTQD 410
QY 589 I-----KQIT-----DKTLNKKIVKDPAR-----ATTVEPILNKDTGEVSEIK 628
DB 411 LIDLANKQVYVIAENIFNIADPKTKAIR-VAKDSABYLGKVSITVTKVNNINSKETEQ-- 467
QY 629 PHRVTVTQNGKEMSGSTV--SEEDFILPVYKGL-----EKGYQPDGMEI-----SGFEK 678
DB 468 FPIELLRVSDMLVVSNSIFGKGEIVIEDSGKITVLANQFNLNGKITYSPKSNANK 527
QY 679 K-----DAGVYNLSKDTFIRPVFKIEKKEE-----NKPTFDVK 716
DB 528 EPIVHDSQGNFRI-VTEGLYTIYVNNLSDKYKAKKNEBKQYNSNNVIDSQKGBFNS 586
QY 717 KDNPOVNSHLSHNR-----EDLREHSHQSDSTKQV 752
DB 587 SKDNQMDKIDKQDNKTEEVNYKIVGGRTEHHNKSKEIVDV 632

RESULT 10
US-09-583-110-3019
Sequence 3019, Application US/09583110
Patent No. 669703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
Pneumoniae for Diagnostics and Therapeutics
FILE REFERENCE: PATHO-07A
CURRENT APPLICATION NUMBER: US/09/583,110
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 3019
LENGTH: 708
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-583-110-3019

Query Match 4.7%; Score 188.5; DB 4; Length 708;
Best Local Similarity 21.4%; Pred. No. 0.00031;
Matches 138; Conservative 92; Mismatches 233; Indels 183; Gaps 33;

QY 223 VSSYGVNVELEGBDLSKRPDLTMSGKLYSDSEKQY---LKDNIILR--KGYALK 277
DB 92 ISSNGTI-----RSNQDLNRT-VES-TVSTVENKSYKEDVLSDRILKKEPFDALS 142
QY 278 VTTN-----PKTDLBNGNGVYSKEDIKIQ-----KANPNLRLSE 315
DB 143 VKDIGAVGDIHDDRQALIDAAAGGLGGGAVTFPGGTIVLKEIVFLKSHTHLELNEK 202

```

QY 316 TTIVADSRNVEDGRSTQSVLMSALDGFNIIRYQVPTFQNMDEKALDKOGLVTDSSKY 375
DB 203 ATI-----LNGINIKNPSIYFMT-----GLTTDGAQY 231
QY 376 LFG-KODKEYTGBDKFNVEAIKEDGMLPDTKPVNL-SM--DKNYNPSKSNXYRNMP 432
DB 232 EWGPTEDISYSGGTIDMNGALNEEG-----TKAKGLPLINSGCAFALGNSNNVTKNVT 285
QY 433 F---YLKGLSDKGGFNMELRVNESVVDNLYIGDHLIDTRPINKLVKQGDIM---D 486
DB 286 FDSYOGHAIQIAGSN-----YVDSRPLGQA-----LPTKDKDQIISKES 339
QY 487 WGMKDYKAGPDKVTDMDQ-----NVYLTQGY---SDLNK---AVGVHYQPLVDNKP 536
DB 330 IQIEPLTRKGFPLALND--DOKSENVITQNSYFGSKDSELYALIGHYQTLSTQPSN 368
QY 537 VNIDPKGNTSIBVADKSVFVFNIDRRNNGPDGEIOGHY-----INGKEYTSFND 588
DB 389 IKLINHPDNMYAGVRFTGFTDVLKGNRFDKKVGESVHYRESGALVNAVSYKNTKD 448
QY 589 I-----KQII-----DKTLNFKI VVDPRP-----NTYKKEFI LKNDGGEVSELK 628
DB 449 LIDLNRQVAVIENIFNIADPKTKAIR-VAKDSAYLGKVS DITVTKVNNNSKETEQ-- 505
QY 629 PHRVVTIONGKEMSSTIV--SEEDPLIPYKGL-----EKGYQPDGMBI-----SGFEK 678
DB 506 FNIELRVSDNLVVSERSIFGKGGIYIBSKKITVLANQFNLGSKTISPKSNANGK 565
QY 679 K-----DAGVYINLSKDTPIKFPFKIEKKEE-----NKPTDVSK 716
DB 566 BPVIRSDGPNFI-VYENGLYKIVTNMLSPDKNEKKEKKEKQYNSNNVIDSNKNGEFS 624
QY 717 KXDNQPVNLSQMLSHRK-----EDLQREHSQKSDSTKV 752
DB 625 SKDNRMKNDKIDNKDNTKEVNYKIVGDGRETENHINKSEIADV 670

RESULT 11
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OR INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Berkreiser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-728-470-10
;
Query Match 4.7%; Score 187.5; DB 2; Length 1529;
Best Local Similarity 20.1%; Pred. No. 0.001;
Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37;
;
QY 9 KFKNL---GNGKESLKKDTTGVGHHHQBENBSIKESSTFIDRNISTRIDFENKDLK-- 63
DB 531 RKNVNSINGTK-----GLKFLANQNPFTHKPDGELNIS-GIVTINQTTKQVYK 580
QY 64 KLIKKEFREYVDFTSETGKMEBYDY-KYDDKANIAYDGDGDLVETKEKLDIYSKIY 122
DB 581 NAKSDSYMNVSSLTNT---VQKFTFIKVDSSG-----NSODLSSRRSPAG 625
QY 123 VLSPSKDGHE-ILKIS-NVSKGAKYV-----GNNTKSIET---KATKDFSKT 169
DB 626 V-----HNGIGGKTNFNIGANAKALFKLQPNAAITDPKKEIPLITNANITATGSDSS 678
QY 170 MTEFLYANIND-----IVDGLAFAGDWRLFVKNDQKKAIEIKIMPEKIKETSEYPVS 224
DB 679 VMFDIHANLTSRAAGIMDSINITGIDPSITSHNRNSNAFEIKDILTNATGNSFS--- 735
QY 225 SYGNVIELGSGDLSKPKDNLTKESGKIYSDSEKQOYLKDNITLAKGVALKVTYTPG 284
DB 736 -----LKQTKDSFYNEYSKHAINSHNLT----- 760
QY 285 KTDMLBGNVYSKEDIKAKIPANPLALSETTIYADSRNVEDGRSTQSVLMSALDGFNI 344
DB 761 ----LGSNVTLLGSSSSSTTGNTININTKANVTLQADTNSNNTGLKAKITLIG----- 809
QY 345 IRYQVTFKRNNDKGEALDKDGNLVTDSSKLVLFGK---DKKEYTGBDKFNVEAIKEDGS 400
DB 810 -----NISVGNLSLTGANANINIGNLSIADSTFKGE----- 841
QY 401 MFLDTKPVNLSDKNYFNPSKSNKIYVRNPEFLRKGISDKGFGNMLRVNESVVDNLY 460
DB 842 -----ASDMLNITGFTNNGTAN-INIKQGVVLQGDINNKGGLN-ITTNASGTOKTI 892
QY 461 IYGDHLIDNTRDNIKNLVK-DGDIMDWG-----MKDYKAN-----G 496
DB 893 INGNL-TNEKGDLNIR-NIRADARITQIGNISQKRGMLTSSDVTNITMOITTAAGYEGG 950
QY 497 FPDKVTMDGNVYLQI-----GY-----SDL-----NAKAVGVHY 526
DB 951 RSDSSBAKANLITITGELADGDLNISGFNKALITAKNSDDLITGASGNADAKKY-- 1008
QY 527 QPLVDNKPPEVNITDPKNTSIBVADKSVFVFNIDRRNNGPD--GGLIOGHYIYNGKEYT 584
DB 1009 --TFDKV-----DSKIST-----DGHVTVINSKVSNSNAGNNSGTGLTISADVT 1056
QY 585 SPNDIQIIDKTNIKIVVADPARNTVKE-FILNKGYSBELKPHRVVTIYONG----- 639
DB 1057 VNNNVTS--KHTINISAA---AGNVTYKGGTTINATTSVY-----VTAONGITKG 1102
QY 640 --KEMSTIVSEDPFI 653
DB 1103 NITSQNVTVATANTIV 1118

```

RESULT 12

```

US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Betkstreaser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-10

Query Match 4.7%; Score 187.5; DB 3; Length 1529;
Best Local Similarity 20.1%; Pred. No. 0.001;
Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37;

QY 9 KKKRL---GNGKESGSIKKDTGVHGHGHEBSIKKSSFTIDRISITIRDPENDLK-- 63
DB 531 KFNVNVSINGTK-----GLKFIANQNPNFTHKFDGLNIS-GIVINOTTKDQVYW 580
QY 64 KLIKKRGAEVVDFTSETKRMEBYDY-KYDDKGNIIAYDGTDEYETKLEIKSKYYG 122
DB 581 NASDSYVNVSSLTNT---VQKFFIKFVDSGS-----NSQDLSSRSRSPAG 625
QY 123 VLSRSKQGHF-IIGKIS-ANSKAKVYV-----GNNYSIER---KATKIDFHSKT 169
DB 626 V-----HFGIGGKTNFNIGANAKALPKLKPNAATDPKBLPTTFNANTATGNSDS 678
QY 170 MTFPLVANIIND-----IYDGLAPAGDMELFYVNDQOKAEIKIRPEKIKETKSEPYVS 224
DB 679 VMFPIHANTLRBAAGINMDSINITGDLPSITSHRNANAFBIKQULTINATGSPFS--- 735
QY 225 SYGVNIEIGEGDLSKNNKPDNI.TPMSGKIYSDSEKQOYLKONITILRRGYALKVTTYNG 284
DB 736 -----LKOTKOSFYNBSKHAINSNNLTI----- 760

```

```

QY 285 KTDMLGKGVSVSKEDIAKIOKANPILRALSETTIYADSRNVEDGRSTQSVLMSLDGFINI 344
DB 761 ----LGNVTLGEGHSSSSITGNINHTKANVTLQADTNSNTGLKKRTLITLG----- 809
QY 345 IRYQVTFKMDKGAIDKQGNLVYDSSKLVLPGR---DKREYGEDKFNVEALKEGDS 400
DB 810 -----NIEBGNLSTGANANIVGNLSIAEDSTPKGE----- 841
QY 401 MLFIDTKPVLNLSMDKNVNPSSKNKYVRNPEFYLRGKISDSDGGFNWELRVNESVVDNYL 460
DB 842 -----ASDNLNITGTFPNTNGTAN-INIKQGVVRLQGDINNGGALN--ITNASTGTOKTI 892
QY 461 IYGDILHINTEDFNILNVK-DGDIIMDG-----MDQYKAN-----G 496
DB 893 INGNL-TWEKEDLNK-NIKADABIQIGANISQKRGNLTISSDKNITNQITIKGVEGG 950
QY 497 PPDKVTMDGNYVLTQF-----GY-----SDL-----NAKAVGVHY 526
DB 951 RSDSSEARNANVLITQKELKLAGDNLISGFNRABITAKRGSDLTIGNAGGNADAKV-- 1008
QY 527 QELYNVNPEVNIIDPKGTSIEYADGKSVFVNIIDKRNNGPD--GEIQGHYIYNGEY 584
DB 1009 --TFDKVR-----DKISIT-----DEHNVTLNSEVKTSGSSAGADNDSGLTISAKV 1056
QY 585 SFNDIKQIITDKTLNIRIYVKDFARNTYVKE-FILNKDTGEVSELKPHRVTVTIQNG--- 639
DB 1057 VNNVTS--HKTINISAA-----AGNVTLKEGTVINATGVSF-----VTAQNGTIG 1102
QY 640 --KEMSSITVSHEDPI 653
DB 1103 NITSQNVTVTATENLV 1118

```

```

RESULT 13
US-09-200-650E-3
; Sequence 3, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Richlin, Deirdre N.
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 3
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-200-650E-3

Query Match 4.6%; Score 186.5; DB 4; Length 930;
Best Local Similarity 20.8%; Pred. No. 0.0006;
Matches 175; Conservative 121; Mismatches 346; Indels 199; Gaps 42;

QY 3 GEIASEKPKNNGKNGKESGIRK-DTGVHGHQ---ENESIKKESFTIDRISITIRDFE 58
DB 56 GELNDSKNEBTAPASNKTKTKVDSKQLKDNQYATADPKYTMDSATVYKERSNNQSFQ 115
QY 59 NKDLKLIKKRGAEVVDFTSETKRMEBYDYKYDDKGNII-AYDGTDEYETKLEIK 117
DB 116 NATNQSTTKTSNVYTNKDSITYSNE-----TDKSNLTQAKDVSTPKTTIKPRTLN 169
QY 118 SKIVGVLSPSKQGHFEILIGKISNVSKAKVYVYVNNYSIRIKATKYDFHSKTMFPLVYAN 177

```

```

Db 170 RMAVNTVAAPQGG-----TNV--NDRKHFNS-----IDIAIDKGVHNOTKTEBFMAT 215
Qy 178 INDI-----VDGLAPAGDMLRPLPKNDQKAKIKIMPEKIKETKSEFPVSSYGVN 229
Db 216 SSDVAKLKANNTIDDSVKGSD--TFTEKYGQYFRPGSVLP--SOTOMLY--NAQGINI 267
Qy 230 IELGEGDLSKNRP-----DNLTVMESGKIYSDSEKQOYLKDNIIIRKGYALVATY 281
Db 268 IAKGIYDSTTTTNTTFTNTVVDYNTVRG-----SFEQVAPAK-----RKNAITDRTKY 316
Qy 282 NRGKDTMLBGCNVYSKEDIQAKQKAPNLRALSETTIYAD---SRNV-----EDGRST 331
Db 317 ---KKEVTLKANDTYSEELI--VDYGNKKAQPLISSSTNYINNEBLSHMTATYVNPRTY 371
Qy 332 OSVLMSALDGF-----NIIRYOVFTPKQNDKGBALIDKQSLVTLVFGKDGKRYTGB 387
Db 372 KQTFPTNLTYGTFNPAKPKFI--YEVDONQFVD--SFTPDYTKL-----KQVT-- 417
Qy 388 DKFNVEALKEGDSMLPITDKFPVNLSDKNYF-----NPSKSNKLYVRNPEFYLRGKI 439
Db 418 DQFDV-IYSNDKKTATVVDLMKGQTSNKGYYIIQVAVAPNNSSTDN-----GKI 464
Qy 440 ----SDKSGFNR--LAVNESVVDNYLIYGDHLIDTRDENIKLANVKG--DIMDKGM 489
Db 465 DYTLDTDKTKYSWSNSYSNNGSSTAN---GPQKKYNLGDYVWEDTNDGKQDANEKGI 520
Qy 490 KQY-----KANPFP-DKVTMDQGNVYLTQYSDDLNAKAVGVH-----YQFLYDNYK 534
Db 521 KGVYVYILKDSNKGKEDRTTIDENGKQYFTGLSN-----GTVSVERSTAGTPTTANVG 574
Qy 535 PEVNIDPKMNTS--IEVADG-----KSVFENINDK--RNNPGDGEIOPQHIIYNGK 581
Db 575 TDDAVDSDELATTGVIKQADNMTLDSGFYKTPKYSGLDYVWYDSNDGKRDSTERSKIGV 634
Qy 582 ERTSNDIKQIIDKTLNITVVKDPAKNTTVEKFIANKOTGEVSEL--KPHRYTVTIQ-- 638
Db 635 KVLTLQNEKEVEIGTT-----ETDNGKTRPDNLDSGKXKVPFKPAGLTOGTNT 684
Qy 639 -----GKEMSSTVSEEDFLPYKGBLEKGYOPDGEIISGFGKQDAGVIMLSKD 690
Db 685 TEDDDADAGGVYVITTHDDF-----TLONGYEEBTSDBSDSDSDSDSDSDSDSD 737
Qy 691 TPIKEVFKKIEBKKEENKPTFDVSKKQNPVNHSQLNESHKEDLQREBSQKSDSTK 750
Db 738 S-----DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 788
Qy 751 D 751
Db 789 D 789

```

```

; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-200-650E-5
Query Match 4.6%; Score 186; DB 4; Length 1315;
Best Local Similarity 20.8%; Pred. No. 0.001;
Matches 188; Conservative 115; Mismatches 328; Indels 274; Gaps 51;

Qy 23 KQTVGVHEHHQRE-----ESIERKSPFTIDRISITJDEPNDL---KK 64
Db 164 KSVNVQPTNEERKQVADTBESTTLLNVKSDAISNDETLVDN--SUSNNENADIIIPKS 222
Qy 65 LKKKF-----REVD--FTSEYKRMERBYDYKDDGN--IIAYDQDTL 106
Db 223 TAPKRLNTMRILAAVQPSSTBAKQNVADITSNTLLTVVD-----ADKNNKI VPAQDYLSTL 277
Qy 107 ERTETKLDRIKS-----KLYGULSPSKQHPFLIGKISNVSKAKATYYGNNYK 154
Db 278 KSOITVDKVKSGDYFTIKRSDTVQVYG--LMPF-----DIKNI--GDIKDPNNGE 324
Qy 155 SJERKATKYDPSKWTPTF--DLXANIINDVGLAFA-----GDMELFYK--DN 198
Db 325 T--IATAKEDTANNLITFTPTVDYRPNVQMGINSITMADPTIYVSKRDVFNNTIGN 382
Qy 199 DQKKAERIKRMEKIKETKSEY-----PYVSSYGN-----VIBLGEEDLSNRK- 241
Db 383 TTKKTANI QYDPYVYNEKNSIGSAFTETVSHVGNKRNPGYKQTIYVNPSENSLTWAKL 442
Qy 242 -----PDNLTKMESGKIYSD--SEKQOYLKDNIIIRKGY-----ALKV 278
Db 443 KVOAVHSSVYNNI--GOINNDVTDIKIYQPKGYTLNKGVDNTEKBLTDVTOYLOK 497
Qy 279 TTYNPKTMLBGNVYSKEDIQAKQKAPNLRALSETTIYADSRNVEDRSTOSVLSA 338
Db 498 ITTGDNNSAVID---PQADSAHYVMVNTKQ-----YMS--ESFTLVQMNTLSS 543
Qy 339 LDGFNIIRYQVTFKKNDRGBA---IDKQGNLY--TDSK-----LVLPK 379
Db 544 TGNKSVSTGNALGFTNNGGAGQGEVYKIGNYWEEDTNKGVQBLERKGVNVTVPFDN 603
Qy 380 DDKETGEKFNVEALIKEDGSMFLI-----DTEPVLNSDKRY--FNPSEKNIKLYVRNPEF 433
Db 604 NNTTKYGE--AVTKEDGSTILPMLPNGDYRVERSNLPKGEYVTPSKG--ANBEL 654
Qy 434 YLRGKISDKGFRNMBELRVNESVVDNYLIYGDHLIDTRDENIKLANVKGQDIMDKGMDYK 493
Db 655 -----DSNGLSSVITVNGK--DN--LSADLGI-----YKPKNML--GDYV--W--EDTN 693
Qy 494 ANGFDPK-----VTKDGNYVLTQYSDDLNAKAVGVHGYQL--YDNVKEPVNI--D 540
Db 694 KNGIQODEKGISGVVTLTKDENGVN--LKVTVTTDAGK-----YKFTDLDMGNYKVEFTT 747
Qy 541 PKGNISIEYADGKSVFENINDKRN-----NGPGEIOBQHIIYNGKRYTSFNDIKQIIDKT 596
Db 748 PEGYFTTYSGSDIEKDSNGLTTTGVINAGADMTLDSGFTKPK---YNLGVNWBOT 803
Qy 804 -----KDKQDSTBKIGSVTTLKN-----ENGVLQ 832
Db 656 VYKGBLEKGYOPDGEIISGF--EGKQDAGVIMLSKDTIKYFKKIEBKKEENKPTFD 713
Qy 833 TTKTMDKGYOFTGLNGYTKVEFPSPGTYTP--QVSGSTDBRIDSNGSTVG 884
Db 714 VSKKQNPVNHSQLNESHKEDLQREBSQKSDSTKD-----YTATVLDKNNISSKST 767
Qy 885 VIKQDNMTIDSGFYKPTYNLGDYVWBDRKNGVQDKDKGISGVTYTLADBRDKYLV 944
Db 768 TNNPN 772
Qy 945 TTIDEN 949

```

```

; SEQUENCE: 1315
; LENGTH: 1315

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:19:57 ; Search time 141 Seconds
(without alignments)
1791.327 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KGEIABSRKYNLGNKRGES.....ATYLDKNKISKSTNNPNK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues
Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/prodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/2/pubppaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubppaa/PCFUS_PUBCOMB.pep:*
7: /cgn2_6/prodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/prodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/prodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/prodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/prodata/2/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/prodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/prodata/2/pubppaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/prodata/2/pubppaa/US10_NEW_PUB.pep:*
18: /cgn2_6/prodata/2/pubppaa/US11_NEW_PUB.pep:*
19: /cgn2_6/prodata/2/pubppaa/US60_NEW_PUB.pep:*
20: /cgn2_6/prodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	US-10-067-385-8	Sequence 8, Appl1
2	4026	100.0	2119	US-09-769-744A-28	Sequence 28, Appl1
3	4026	100.0	2140	US-10-282-122A-73670	Sequence 73670, A
4	4026	100.0	2140	US-10-472-928-1180	Sequence 1180, Ap
5	615	15.3	117	US-09-765-472-68	Sequence 68, Appl1
6	227	5.6	1196	US-10-282-122A-52737	Sequence 52737, A
7	222	5.5	861	US-09-820-843A-109	Sequence 109, Appl1
8	218	5.4	1639	US-10-087-464-10	Sequence 10, Appl1
9	212	5.3	1141	US-10-282-122A-70251	Sequence 70251, A
10	207.5	5.2	4688	US-10-282-122A-76865	Sequence 76865, A
11	207	5.1	1191	US-10-282-122A-52048	Sequence 52048, A
12	203	5.0	1166	US-10-744-616-7	Sequence 7, Appl1
13	203	5.0	1881	US-10-032-585-7646	Sequence 7646, Ap

14	200.5	5.0	1786	9	US-09-742-096-3	Sequence 3, Appl1
15	200.5	5.0	1787	15	US-10-415-253-2	Sequence 2, Appl1
16	196.5	4.9	5184	16	US-10-282-122A-53254	Sequence 53254, A
17	196.5	4.9	5176	16	US-10-437-963-150986	Sequence 150986, A
18	195	4.8	1849	15	US-10-637-544-2	Sequence 2, Appl1
19	194	4.8	1178	15	US-10-282-122A-52434	Sequence 52434, A
20	194	4.8	1957	15	US-10-369-493-2070	Sequence 2070, Ap
21	193	4.8	903	15	US-10-282-122A-52328	Sequence 52328, A
22	193	4.8	909	15	US-10-282-122A-52109	Sequence 52109, A
23	192.5	4.8	708	17	US-10-472-928-3792	Sequence 3792, Ap
24	191.5	4.8	1156	15	US-10-369-493-43	Sequence 43, Appl1
25	190	4.7	1788	15	US-10-282-122A-43832	Sequence 43832, A
26	190	4.7	1089	15	US-10-282-122A-46664	Sequence 46664, A
27	189.5	4.7	841	9	US-09-815-242-5779	Sequence 5779, Ap
28	189.5	4.7	841	9	US-09-815-242-12751	Sequence 12751, A
29	189.5	4.7	1143	15	US-10-369-493-11081	Sequence 11081, A
30	189	4.7	1313	15	US-10-282-122A-76863	Sequence 76863, A
31	188	4.7	943	17	US-10-741-849-7309	Sequence 7309, Ap
32	187	4.6	1103	15	US-10-282-122A-76866	Sequence 76866, A
33	187	4.6	2265	15	US-10-282-122A-45123	Sequence 45123, A
34	186.5	4.6	930	17	US-10-744-616-3	Sequence 3, Appl1
35	186.5	4.6	1790	15	US-10-369-493-1586	Sequence 1586, Ap
36	186.5	4.6	1903	16	US-10-766-993-3	Sequence 3, Appl1
37	186	4.6	1315	17	US-10-744-616-5	Sequence 5, Appl1
38	185	4.6	1385	15	US-10-282-122A-44324	Sequence 44324, A
39	185	4.6	3225	16	US-10-408-765A-254	Sequence 254, App
40	185	4.6	5005	15	US-10-282-122A-76871	Sequence 76871, A
41	184.5	4.6	1946	15	US-10-282-122A-62947	Sequence 62947, A
42	184	4.6	1875	15	US-10-369-493-22285	Sequence 22285, A
43	183	4.5	2402	16	US-10-661-809-20	Sequence 20, Appl1
44	182.5	4.5	882	15	US-10-282-122A-53247	Sequence 53247, A
45	182.5	4.5	923	16	US-10-771-931-31	Sequence 31, Appl1

ALIGNMENTS

RESULT 1									
US-10-067-385-8									
Sequence 8, Application US/10067385									
Publication No. US20020110562A1									
GENERAL INFORMATION:									
APPLICANT: Adamou, John									
TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines									
FILE REFERENCE: 469201-589									
CURRENT APPLICATION NUMBER: US/10/067,385									
CURRENT FILING DATE: 2002-02-05									
PRIOR APPLICATION NUMBER: US/09/590,991									
PRIOR FILING DATE: 2000-06-09									
PRIOR APPLICATION NUMBER: US/60/138,453									
PRIOR FILING DATE: 1999-06-10									
NUMBER OF SEQ ID NOS: 8									
SOFTWARE: PatentIn Ver. 2.0									
SEQ ID NO 8									
LENGTH: 773									
TYPE: PRT									
ORGANISM: Streptococcus pneumoniae									
US-10-067-385-8									
Query Match									
Best Local Similarity 100.0%; Score 4026; DB 13; Length 773;									
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	KGEIABSRKYNLGNKRGESIKKQTTVEHHHONKESISFKSSFTIDRNISTRDPNK	60						
DB	1	KGEIABSRKYNLGNKRGESIKKQTTVEHHHONKESISFKSSFTIDRNISTRDPNK	60						
QY	61	DLKKLKKKRRVDDPTSEGRKMEYDYKDDKGNITAYDDGTDLEKTEKLEISKI	120						
DB	61	DLKKLKKKRRVDDPTSEGRKMEYDYKDDKGNITAYDDGTDLEKTEKLEISKI	120						
QY	121	YGVLSPSKDGHEFTLIGKISNYSKNKAVYYGNMYKSIIEIKATKYPFHSKTMFDLYANIND	180						


```
Db      ||||| 121 YGVLSPSKDGFELLGKISNVSNAKAVYGNKYKSEIRATKTDFFSKMTTPDLANIND 180
Qy      ||||| 181 IVDGLAFAGDMELFVNDNQKAEIKIRMPBKIKETKSEYPIVSSYGVNIELGEGDLSKN 240
Db      ||||| 181 IVDGLAFAGDMELFVNDNQKAEIKIRMPBKIKETKSEYPIVSSYGVNIELGEGDLSKN 240
Qy      ||||| 241 KPDNLTMSGKTIYSDEKQOYLKONILIRKGYALKVTTNPCKTDLBGNGVYSKEDI 300
Db      ||||| 241 KPDNLTMSGKTIYSDEKQOYLKONILIRKGYALKVTTNPCKTDLBGNGVYSKEDI 300
Qy      ||||| 301 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYOVTFPGNDKGEA 360
Db      ||||| 301 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYOVTFPGNDKGEA 360
Qy      ||||| 361 IDKXGNLVTDSSKLVLFKGDOKXYTGEBDFNVEALIKEDGSMLEIDTKPVNLSMDKNYFNP 420
Db      ||||| 361 IDKXGNLVTDSSKLVLFKGDOKXYTGEBDFNVEALIKEDGSMLEIDTKPVNLSMDKNYFNP 420
Qy      ||||| 421 SRSNKTIVYRNPEFYLRGKISDKGFWMELRVNESVVDNLTLYGDLHIIDNTRDPNLIKLVK 480
Db      ||||| 421 SRSNKTIVYRNPEFYLRGKISDKGFWMELRVNESVVDNLTLYGDLHIIDNTRDPNLIKLVK 480
Qy      ||||| 481 DDDIMDMGKDYKANGFPDKVTDMGAVNLTQTSYSDLNKAVGVHYQFLYDNVKEPVNID 540
Db      ||||| 481 DDDIMDMGKDYKANGFPDKVTDMGAVNLTQTSYSDLNKAVGVHYQFLYDNVKEPVNID 540
Qy      ||||| 541 PKGNTSIBYADGKSVFVFNINDKRNNGPDGEIOGHYIYNGKETSFNDIKQIIDKTLNIK 600
Db      ||||| 541 PKGNTSIBYADGKSVFVFNINDKRNNGPDGEIOGHYIYNGKETSFNDIKQIIDKTLNIK 600
Qy      ||||| 601 IYVNDPARNVTYKSEFLINMDTGEVSELKPHRVTVTIQNGKEMSTIVSEBDFILPYKGE 660
Db      ||||| 601 IYVNDPARNVTYKSEFLINMDTGEVSELKPHRVTVTIQNGKEMSTIVSEBDFILPYKGE 660
Qy      ||||| 661 LEKGYPDGMELISGFGKDGAGVYNLSKOTFIKPVFKKIBEKKEENKPTFVYSKXKN 720
Db      ||||| 661 LEKGYPDGMELISGFGKDGAGVYNLSKOTFIKPVFKKIBEKKEENKPTFVYSKXKN 720
Qy      ||||| 721 POWNHSQLNESHKEDLQREBHSQKSDSTKDVTATVLDKNNTSSKSTNNPNK 773
Db      ||||| 721 POWNHSQLNESHKEDLQREBHSQKSDSTKDVTATVLDKNNTSSKSTNNPNK 773

RESULT 2
US-09-769-744A-28
; Sequence 28, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard W
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hambley, Sean B
; APPLICANT: Hambro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 4026; DB 10; Length 2119;
Best Local Similarity 100.0%; Pred. No. 5.8e-216;
```

```
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      ||||| 1 KLGIAESFYKNLGNKESGLKDDTTGVHHHQBNEESIKKSSPTIDNISTIRDPENK 60
Db      ||||| 1313 KLGIAESFYKNLGNKESGLKDDTTGVHHHQBNEESIKKSSPTIDNISTIRDPENK 1372
Qy      ||||| 61 DLKLIKKRREVDPTSEFGKMEFYDKYDKKNIIVDDGTDLEFTEBKDEIKSK 120
Db      ||||| 1373 DLKLIKKRREVDPTSEFGKMEFYDKYDKKNIIVDDGTDLEFTEBKDEIKSK 1432
Qy      ||||| 121 YGVLSPSKDGFELLGKISNVSNAKAVYGNKYKSEIRATKTDFFSKMTTPDLANIND 180
Db      ||||| 1433 YGVLSPSKDGFELLGKISNVSNAKAVYGNKYKSEIRATKTDFFSKMTTPDLANIND 1492
Qy      ||||| 181 IVDGLAFAGDMELFVNDNQKAEIKIRMPBKIKETKSEYPIVSSYGVNIELGEGDLSKN 240
Db      ||||| 1493 IVDGLAFAGDMELFVNDNQKAEIKIRMPBKIKETKSEYPIVSSYGVNIELGEGDLSKN 1552
Qy      ||||| 241 KPDNLTMSGKTIYSDEKQOYLKONILIRKGYALKVTTNPCKTDLBGNGVYSKEDI 300
Db      ||||| 1553 KPDNLTMSGKTIYSDEKQOYLKONILIRKGYALKVTTNPCKTDLBGNGVYSKEDI 1612
Qy      ||||| 301 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYOVTFPGNDKGEA 360
Db      ||||| 1613 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYOVTFPGNDKGEA 1672
Qy      ||||| 361 IDKXGNLVTDSSKLVLFKGDOKXYTGEBDFNVEALIKEDGSMLEIDTKPVNLSMDKNYFNP 420
Db      ||||| 1673 IDKXGNLVTDSSKLVLFKGDOKXYTGEBDFNVEALIKEDGSMLEIDTKPVNLSMDKNYFNP 1732
Qy      ||||| 421 SRSNKTIVYRNPEFYLRGKISDKGFWMELRVNESVVDNLTLYGDLHIIDNTRDPNLIKLVK 480
Db      ||||| 1733 SRSNKTIVYRNPEFYLRGKISDKGFWMELRVNESVVDNLTLYGDLHIIDNTRDPNLIKLVK 1792
Qy      ||||| 481 DDDIMDMGKDYKANGFPDKVTDMGAVNLTQTSYSDLNKAVGVHYQFLYDNVKEPVNID 540
Db      ||||| 1793 DDDIMDMGKDYKANGFPDKVTDMGAVNLTQTSYSDLNKAVGVHYQFLYDNVKEPVNID 1852
Qy      ||||| 541 PKGNTSIBYADGKSVFVFNINDKRNNGPDGEIOGHYIYNGKETSFNDIKQIIDKTLNIK 600
Db      ||||| 1853 PKGNTSIBYADGKSVFVFNINDKRNNGPDGEIOGHYIYNGKETSFNDIKQIIDKTLNIK 1912
Qy      ||||| 601 IYVNDPARNVTYKSEFLINMDTGEVSELKPHRVTVTIQNGKEMSTIVSEBDFILPYKGE 660
Db      ||||| 1913 IYVNDPARNVTYKSEFLINMDTGEVSELKPHRVTVTIQNGKEMSTIVSEBDFILPYKGE 1972
Qy      ||||| 661 LEKGYPDGMELISGFGKDGAGVYNLSKOTFIKPVFKKIBEKKEENKPTFVYSKXKN 720
Db      ||||| 1973 LEKGYPDGMELISGFGKDGAGVYNLSKOTFIKPVFKKIBEKKEENKPTFVYSKXKN 2032
Qy      ||||| 721 POWNHSQLNESHKEDLQREBHSQKSDSTKDVTATVLDKNNTSSKSTNNPNK 773
Db      ||||| 2033 POWNHSQLNESHKEDLQREBHSQKSDSTKDVTATVLDKNNTSSKSTNNPNK 2085

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```

FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282.1222A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patentin version 3.1
SEQ ID NO 73670
LENGTH: 2140
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-282-1222A-73670

```

```

Query Match 100.0%; Score 4026; DB 15; Length 2140;
Best Local Similarity 100.0%; Pred. No. 5.9e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGRIASKEFKNLGNGKSGSLKDDTGVVHHHQNBSISEKSSFTIDRNISTIRDFENK 60
DB 1334 KLGRIASKEFKNLGNGKSGSLKDDTGVVHHHQNBSISEKSSFTIDRNISTIRDFENK 1393
QY 61 DLKKLKKKFRVDDFTSETKRMEYDYKTDKGNIIAYDDGTDLEYTEKLDSEISKI 120
DB 1394 DLKKLKKKFRVDDFTSETKRMEYDYKTDKGNIIAYDDGTDLEYTEKLDSEISKI 1453
QY 121 YGVLSPSKDGHEFLIGKISNVSKNAKYGGNNYSIEIKATKYPHSHKTMFPDLAYANIND 180
DB 1454 YGVLSPSKDGHEFLIGKISNVSKNAKYGGNNYSIEIKATKYPHSHKTMFPDLAYANIND 1513
QY 181 IYDGLAFAGDMRLFVKONDQKAEIKIRMPERIKETSESEYEVYSSYGAVIELGEGDLSKN 240
DB 1514 IYDGLAFAGDMRLFVKONDQKAEIKIRMPERIKETSESEYEVYSSYGAVIELGEGDLSKN 1573
QY 241 KPDNLITWMSGKISYDSBKQOYLLKDNIIILKGYALKVTYTPGKTMDLBSNGVYSKEDI 300
DB 1574 KPDNLITWMSGKISYDSBKQOYLLKDNIIILKGYALKVTYTPGKTMDLBSNGVYSKEDI 1633
QY 301 AKIOKANPNLALSETTITVADSRNVEDGRSTQSVLMSALDGFNIIIRYOVFTFKANDKGEA 360
DB 1634 AKIOKANPNLALSETTITVADSRNVEDGRSTQSVLMSALDGFNIIIRYOVFTFKANDKGEA 1693
QY 361 IDKGNLVTSSSKVLVFGKDDKEYTGSDKFNVEAIKEDGSMFLPDTYKPVNLSMDKNYFNP 420
DB 1694 IDKGNLVTSSSKVLVFGKDDKEYTGSDKFNVEAIKEDGSMFLPDTYKPVNLSMDKNYFNP 1753
QY 421 SKSNKIVYRNPDEPLRGKISDYGKGMWELRVNESVDNNTLIYDGLHIDNTRDPNLIKLVNK 480
DB 1754 SKSNKIVYRNPDEPLRGKISDYGKGMWELRVNESVDNNTLIYDGLHIDNTRDPNLIKLVNK 1813
QY 481 DGDIMDGMKMDYKANGFPDVKTYDDMGVNYLQTYGSDLNAAKAVGVHYOFLYDNPVPEVNID 540
DB 1814 DGDIMDGMKMDYKANGFPDVKTYDDMGVNYLQTYGSDLNAAKAVGVHYOFLYDNPVPEVNID 1873
QY 541 PKGNTSIEVADGKSVFNINDKRNNGFDGELQBOHIIYINGKEYSFNDIKOIIDTKTLNIK 600

```

```

DB 1874 PKGNTSIEVADGKSVFNINDKRNNGFDGELQBOHIIYINGKEYSFNDIKOIIDTKTLNIK 1933
QY 601 IYVQDFARNTTVEKFIINAKDTGEVSELKPHRYVTTIIONGKMSSTIYSEEDPILPYNGE 660
DB 1934 IYVQDFARNTTVEKFIINAKDTGEVSELKPHRYVTTIIONGKMSSTIYSEEDPILPYNGE 1993
QY 661 LEKGYPDGMWISGFEKGDAGVYINLSKQFFIPVFPKIEKKBBBKPPFDVSKKDN 720
DB 1994 LEKGYPDGMWISGFEKGDAGVYINLSKQFFIPVFPKIEKKBBBKPPFDVSKKDN 2053
QY 721 POWNHSQLNSHREKDLQREHSQKSDSTQVATVLDKNNISSKSTNNPNK 773
DB 2054 POWNHSQLNSHREKDLQREHSQKSDSTQVATVLDKNNISSKSTNNPNK 2106

```

```

RESULT 4
US-10-472-928-1180
Sequence 1180, Application US/10472928
Publication No. US20050020813A1
GENERAL INFORMATION:
APPLICANT: CHIRON SPA
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
CURRENT APPLICATION NUMBER: US/10/472,928
FILE REFERENCE: P026926MO
PRIOR FILING DATE: 2003-09-26
PRIOR APPLICATION NUMBER: GB-0107658.7
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 4979
SOFTWARE: SeqMan99, version 1.03
SEQ ID NO 1180
LENGTH: 2140
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
FEATURE:
OTHER INFORMATION: serine protease, subtilase family
OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LpxTG)
OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.B+01)
US-10-472-928-1180

```

```

Query Match 100.0%; Score 4026; DB 17; Length 2140;
Best Local Similarity 100.0%; Pred. No. 5.9e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGRIASKEFKNLGNGKSGSLKDDTGVVHHHQNBSISEKSSFTIDRNISTIRDFENK 60
DB 1334 KLGRIASKEFKNLGNGKSGSLKDDTGVVHHHQNBSISEKSSFTIDRNISTIRDFENK 1393
QY 61 DLKKLKKKFRVDDFTSETKRMEYDYKTDKGNIIAYDDGTDLEYTEKLDSEISKI 120
DB 1394 DLKKLKKKFRVDDFTSETKRMEYDYKTDKGNIIAYDDGTDLEYTEKLDSEISKI 1453
QY 121 YGVLSPSKDGHEFLIGKISNVSKNAKYGGNNYSIEIKATKYPHSHKTMFPDLAYANIND 180
DB 1454 YGVLSPSKDGHEFLIGKISNVSKNAKYGGNNYSIEIKATKYPHSHKTMFPDLAYANIND 1513
QY 181 IYDGLAFAGDMRLFVKONDQKAEIKIRMPERIKETSESEYEVYSSYGAVIELGEGDLSKN 240
DB 1514 IYDGLAFAGDMRLFVKONDQKAEIKIRMPERIKETSESEYEVYSSYGAVIELGEGDLSKN 1573
QY 241 KPDNLITWMSGKISYDSBKQOYLLKDNIIILKGYALKVTYTPGKTMDLBSNGVYSKEDI 300
DB 1574 KPDNLITWMSGKISYDSBKQOYLLKDNIIILKGYALKVTYTPGKTMDLBSNGVYSKEDI 1633
QY 301 AKIOKANPNLALSETTITVADSRNVEDGRSTQSVLMSALDGFNIIIRYOVFTFKANDKGEA 360
DB 1634 AKIOKANPNLALSETTITVADSRNVEDGRSTQSVLMSALDGFNIIIRYOVFTFKANDKGEA 1693
QY 361 IDKGNLVTSSSKVLVFGKDDKEYTGSDKFNVEAIKEDGSMFLPDTYKPVNLSMDKNYFNP 420
DB 1694 IDKGNLVTSSSKVLVFGKDDKEYTGSDKFNVEAIKEDGSMFLPDTYKPVNLSMDKNYFNP 1753

```



```

Db      240 IEDKLNK-NISSNMLSGENIDINKENKSOYKEIISKF-----NEKLELTK--NNSR 288
Qy      144 NAKVYNNYKSIIEIKATKYDFHSKTMTPDLVANIINDIYDGLAFAGDMRLFYDNDQKKA 203
Db      289 NKEEYVYNDKNDIE-----NENMLKEKIKNLKD-----NIKVENLTKTN 331
Qy      204 EIKIR-----MPEKIKETKSEYRYV----- 223
Db      332 BEELQYLYKEKALEKKINDLKEEESYLRBEIGKEDNVNNFNKLEKEKELKILKSEB 391
Qy      224 -----SSGVNIELEGEDL--SKNKPDLTYMESKIVS-----DSEKQOYLK 265
Db      392 IELTSTSLKRNELISVGENEIVLNEKLDNINKNSCDSYISSININIKTKEDIEKEINIK 451
Qy      266 DNIIIL-----RKGYALKVTTYNPGKTMLEGNVYSKED-----IAKIQK--AN 307
Db      452 ENILLLENMLKENSRLNISLKIISLNNEK--KLKKNAAVSLRLEANNYMLSLNEKHYEGYN 510
Qy      308 PNILALSETTYIYADSRNVEGKSTQSVLMSALDGFNIIRYQVTPFKRNDKGAIDKQNL 367
Db      511 RSYVYTLMEHYSGKVDNIKG-----CEVLDGDIHKKELEIAMEIALGAI--SNV 560
Qy      368 VPDSSKLVLFKGDKEVYTGEDKFNVEALKEKGSWLPIDTKPVNLSMDKNYENP--SKSN 424
Db      561 IIEDE-----NKAKLILNYLKCK-----SLGRATFLPIITIQGR 594
Qy      425 KIYVRNPEFYLRGKISDKGPFNMLRNVESVVDNYLLYGDLIHIDTDFNI--KLAVYDG 482
Db      595 KAKINN-----VREDEGF--LGIASDLIDY--DVKEFNIIIDYVIGRLVAKD- 637
Qy      483 DINDWKGKDYKANGFPKATYDMDGNVYLYQGYSDLMANKAVNHQPLVDNKPENVIDPK 542
Db      638 --MDSALAKIAKKNYSFKITVLRGEV-----INPG 665
Qy      543 GNT-----SIEYADGKSVFVFNNDKKNNGFDGEIOQHITYNGKEYTSFNDIKQIDKTLMI 599
Db      666 GSLTGSISIKRAGSSI--ISRRK-----RIEB-----TKCELBEETKN- 700
Qy      600 KIIVKDPARTTYKEFLTNKDTGSEVSLKPHRVTVTTONGKMSSTVSEEDFLPPYK 659
Db      701 -----TIEEFMKN-----ILENKVKIKIT--LDBENLMI--KD 728
Qy      660 ELKGVQFDGMEISGFEGKDOA-----GYVINLSKD--TFIKVPFKKIEEKEBENK 709
Db      729 EI-----YNNIEIYFKMGKLAIAIKEDTERLRSSLSNISBEIKTLTKDKIOLSEINISOK 784
Qy      710 PTFDVSKKKQNPQVNHSQLNESHKEDLOREHSHQKSDSTK---DYATVLDKNNISSK 765
Db      785 QLEBELKAKDO--LNNHDIKEC--EDFLQNEEENNVKNIKDLIEYKIEKALDEMLVSIK 839
Qy      766 -----STYNNPK 773
Db      840 KELYSMDNTNITNLNENK 857

RESULT 7
US-09-820-843A-109
; Sequence 109, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820, 843A
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 109
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:

```

```

; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: g1|4493994
US-09-820-843A-109

Query Match      5.5%; Score 222; DB 10; Length 861;
Best Local Similarity 19.3%; Pred. No. 0.00037;
Matches 162; Conservative 136; Mismatches 324; Indels 216; Gaps 37;

Qy      2 LGELAESKFPNLAG-----KESGLKDDTGVSEHH--QENESIKESSTFI 47
Db      154 LNNINGKVFKNLDNCGNLPNNLYIDKEG--KMLTGKEHYNAASNEVYNNNNKNTNYY 211
Qy      48 DRNISTTRDPENKDL-----KLIKKKREYVDFTSBTGRMEBYDKYDQKGI 98
Db      212 NNNNSYNNNNFCNNNYNDNNYNNNSNNKMGKNTYERSILYK--KEHMDVDEYNNKNTIR 268
Qy      99 AYD-----DGTDEYBTEKLDIEKSKLYGVLSPEKDGHFBI--LGIISNVSKNAKVYGN 151
Db      269 KNDSEKTVNDPPLHY-----SKNNYDIFLGDIKKAKONKCKGN 310
Qy      152 NYSIIRIKATKYDFHSKTMTPDLVANIINDIYDG--LAFAGDMRLFYDNDQKALICIR 208
Db      311 KYNNM-----HDNNSN-----NSNNVLANNNNSNSNNYNNIFPDNDENL----- 351
Qy      209 MPEKIKETKSEYRYVSSGVNIELEG--DLSK--NKPDNLTYMESKIVSDSEKQOYL 264
Db      352 -----TKSNFAPKFNKNNNNNNVNTDIIKYLNK-----NSQGHSDGKNNNNNN 396
Qy      265 KNTILRKGYALKVTTYNPGKTMLEGNVYSKEDIKIQANPR--LALSETTYIAD 321
Db      397 GNNII-----NNNSNNKNNI FQGSRYEAVMYVYNNNNNNIISNNKREASFTVD 447
Qy      322 SRNVEDGRSTQSVLMSALDGFNIIRYQVTPFKRNDK--EALDKQNLVTDSSKLVLEKGD 380
Db      448 NINTNSGRBEKISNTYAR--LIMKOISMIKERNKGLDVLEKQNT--FGFL 494
Qy      381 DKEYTGDKFVVEALIKDGSWLPIDTKPVNLSMDKNYENPESKNKIYVRNPEFYL----- 435
Db      495 DNNYQNTYGSNN-----NSLSEKNNM--KENDIYSEKASKRIMDIFR 533
Qy      436 -----RGKISDKGPFNMLRNVESVVDNYLLYGDLIHIDTDFNIKLVNKGADINDWKGKD 491
Db      534 TLNNGVLVSQ-----SLTVNQSIVLNNNNNNYNNNSNNRKNQNNNNNNNNNNM----- 582
Qy      492 YKANGFPDKVTDMDGNVYLYQGYSDLNK--AVGVHYQPLVDNKPENVIDPKGVTSIE 548
Db      583 -----NNNNNSNNNNNNNNNNNNYKNNHKKYHSMNVYTKCIF--INNSYNNNDGNNNSNS 633
Qy      549 YADGKSVFVFNNDKKNNGFDGEIOQHITYNGKEYTSFNDIKQIDK----- 595
Db      634 NSNNNVHEYYNNNNKKN--FKNNKINNTHMLPNNKNNNNNNNNYNNINNNLNMENPPSL 691
Qy      596 TLNLIKIVKDPARN-----TVKEPILNKDTGSEVSLKPHRVTVTTONGKMSSTVSEED 651
Db      692 SPNNSDINRNAQGNINITPIINSILRLD--NEVDVHNSISEINQAK--VSNVLDLSIK 748
Qy      652 FILPVYKGELEKGYQFDGMEISGFEGKDOAGVNLSDDTIK-----PVFKKIEBK 704
Db      749 SILKASKQGNNNNNIIP--KNFNNNN-----NNNSKILNYSQOYTSBHQOQOQH 799
Qy      705 EENKPTFVDSKKQNPQVNHSQLNESHKEDLOREHSHQKSDSTK---SDSTKDYATVLDK 759
Db      800 QQQQQQ--QQQQTILQYQINSLHNDPKKGGPKKERRPKYPRBDGTTNETMETVREK 855

RESULT 8
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chientli, Achar

```

```

; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuecong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Patent version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; US-10-087-464-10

Query Match      5.4%; Score 218; DB 14; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.0014;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy      23 KDTTG-VEHHQEMESIK-----EKSFTIDRNISTIRDPENKDKLKKKREYVD 75
Db      232 KDNVGMEDYIKKKKTIENINELIESKKTIDKKRNATKEB-----KKKLYQAY 283
Qy      76 FTSEGRMEERYDYKDKGNIAVDGTDLEYETBEKLDISKIVGVSPEKQHPIL 135
Db      284 DLSTYKQLEB-----AHMLISV-----LE--KRIDTLK-----NENIKEL 319
Qy      136 GISNVSNAKAYVYNNYKSIEIKATYDF-----SKTVPDLANINDIVGLAFA 188
Db      320 DKINIKRPPRANSNTPTTLDKQKIBEHKEIKETAKTIKYNIDSLFTPL----- 373
Qy      189 GDMRLFVKNDOKKAEIKIMEPEKIKETSEYP-----YVSSYGVN-----TELGE-GDLSEK 239
Db      374 -ELAYTLAEKRN-KNIDISAKVETKESTEPENYGVYPLSYNDINNALINELINSFGDL- 430
Qy      240 NKPNULTMESGKIYDSEKQYL--LKDNIL-----RKGVALKVTTYNPKTKML 289
Db      431 -NPDYTKESPKNITTDNERKKFPIBEIKETIKBEKKKIESDKSYEDSKSLN---DIT 485
Qy      290 EG-----NGVYSKE-----DIAKLOKA-----NPNRLALETITYADSRNVEDGRSTQSV 334
Db      486 KEYEKTLAIBEYDSKRNKNIDLTNPEKMMKGRYSYVEKLTHTNFPAYSNLSGNLEKLTJK 545
Qy      335 LMSALDGF---NII-----RYQVTFPKANDKGAIDQGNLVTDSSKLVFGKDKRYT 385
Db      546 ALKTMEDYSLRNIVVEKEIKAYKKNLISKIENIEITLVE--NIKQEBQLF-----EKIT 598
Qy      386 GEDKFNVAIKEDSGMLFIDTPRVNLSMDKVFYNSKSNKIYVRNPEFLRGKISDKGAF 445
Db      599 KDNKPEDEKILEVSDIVAYQVQKYL--ANKIDELKKTQTLKQVE--LKNHIVPNSTY 653
Qy      446 NMELRVNESVVDNYLYGDLHIDNTRDFNIXL-----NVK----- 480
Db      654 KOENKQOB---PYVLIYVAKKEIDKLKVFMPKVESLINERKKNIKTREGSDNSEPSTGBEI 709
Qy      481 -----DGDIDMGKQDYK----- 501
Db      710 TGOATTKRGGQAGSALDEDSVQAQAOEQKQAPVPVPVPEAKAVPTPPAPVNNKTRNV 769
Qy      502 TDMGNAVLYQGYSDINA-----KAVGVHYQFLYDNVPRVAVNIDKQNTSIEYADGKSV 556
Db      770 SKLD---YLEKLYBEFLNTSYIYCHKIYVSHSTPMNEKIKQYKTIYBESKLSGCDPLDL 826
Qy      557 FNIN-----DKRNGPFG---EIQEQHIYNGKSYTSFNDIKOIIDFTLNIKIV 603
Db      827 FNIGNINLVNYSMPFSLNNSLSQLEFMEIYEKMYCNLTKLNDKNDIKMLLEAKKAVSTSV 886
Qy      604 KDPANNTVKEFLINKQNGEYSELKPRHVTVTYIOWGKMSSTIYSEEDFILPYVYGELEK 663
Db      887 KTLSSSQPLSLTPTQDPRVSAANDTSHSTVNLNLSLGFENILS-----LGKKN 937

```

```

Qy      664 GYFDGMEISGPBEK-----KQAGVYINLSKDTFIRPVKRIEKKREBNKPTF--D 713
Db      938 IYQ-----ELIQKSSBNFYKTLKSDPTFYNBSFTNPFVSKADNINLSDESKRKLLED 993
Qy      714 VEKKDNPQVNSQLNESHREKEDLQREBHSQSDSTKDYATVADKNINSSK-STYNNP 771
Db      994 INKLKKTQLSLPDLNRYKTLKLERLPDKKQYVGKTKMQIKKLTLEKQLESKLNSLNP 1052

RESULT 9
US-10-282-122A-70251
; Sequence 70251, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELATRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 70251
; LENGTH: 1141
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-10-282-122A-70251

Query Match      5.3%; Score 212; DB 15; Length 1141;
Best Local Similarity 20.9%; Pred. No. 0.0019;
Matches 182; Conservative 116; Mismatches 336; Indels 238; Gaps 44;

Qy      33 QEMESISGE--KSSFTIRANISTIRDPENKDKLKKKFR-----EYDDFTS 78
Db      221 KNNPEKAEKELRVNNTNTRSTKPVATAPTSVAPKELNKKMPRAVAQPAVASNNVNDLIT 280
Qy      79 ETGRMEERYDYKDKGNIAVDGTDLEYETE-----KLDEIKSKIYGVLS 125
Db      281 VTKQYIK-----VGQKDNVAALHDKQDIEYDTEFLINDKVKRGDMTINIDGNVIPSULT 336
Qy      126 PSKD-----GHEFLIGKISNVSNAK---VYGGNNYKSIEIKATYDFHSKTMTDL 174

```



```

Qy 522 VGHVQFLYDNVKEVNIDPFGNTSIEYADGKSVFNINDK-----RNGFDEIGQOH 575
Db 624 IDKEDLNTLDS-HBPFYDNDGNLEIH-----TOLANDNDLDKOKALNNANVKGIVVQO- 677
Qy 576 IYINGKEYSNDIKQIIDKTINIKIYVCDPABNTVRE--FILNKGGEVSELPKHKVT 633
Db 678 ---DJEH---EIDVSIQANGKVIIPKMLANDPTENITTLK----- 716
Qy 634 VTIONGKEMSTIVSE-----DPLPYKGELEKGYOPGWEIS----- 673
Db 717 VLKQNNQPNIDLISEQSGDNHISFKPKPTTAKTEN--DYEISFSPNPLANKKIK 773
Qy 674 -----GFEKGDAGVYNLSQDTIKPVPKIKIEKGEENKPTDV 714
Db 774 LTFPTDNNNTKTVESIGLDGA-----IFKTSDAIPADPHKTYTLTKIEADNKKVANI 829
Qy 715 SKKDNPQVNHSSOLNESHREKEDLOREHSSOKSDSTKDVATVLDKGN----- 761
Db 830 DEISPLDRIVAKQKNGNANADN--KHFKIPDQKKDLJAVYKDKNNNEIHVPIKTDDKG 887
Qy 762 ---ISSKSTNNPNK 773
Db 888 KVIYVNNMLPDPNK 902

```

RESULT 11

```

US-10-282-122A-52048
: Sequence 52048, Application US/10282122A
: Publication No. US20040029129A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari
: APPLICANT: Zyckind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: EILTRA.034A
: CURRENT APPLICATION NUMBER: US/10/282,122A
: CURRENT FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/230,335
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: 60/230,347
: PRIOR FILING DATE: 2000-09-09
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/267,636
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 78614
: SOFTWARE: Patent version 3.1
: SEQ ID NO 52048
: LENGTH: 1191
: TYPE: PRT
: ORGANISM: Clostridium acetobutylicum

```

US-10-282-122A-52048

Query Match 5.1%; Score 207; DB 15; Length 1191;
 Best Local Similarity 19.6%; Pred. No. 0.0038;
 Matches 178; Conservative 139; Mismatches 275; Indels 314; Gaps 43;

```

Qy 1 KUGELAEKPYKNGKNGK-----SLKQTYVHHHQBBSBI 39
Db 120 RLKDIQF-LFMDYTGKGYSIIGQKIEAVLSGPEERRALIEAGIVKPKRKVD- 177
Qy 40 KKSSEFIDRINISTRD-----FKNQDLKLIKKEFREYVD--FTSETGK--RMEB 86
Db 178 -EKLEANTQNLVRINDLRTYERLEPLIESEAKFVELSDLKTREINTIYISDN 236
Qy 87 YDYKDDKGNITL-----YDGT-----DLEYTEKLDI-----KXIIYGVLSF 126
Db 237 IDYRINDLKQWADLKLSIDENVDKESISLEKVAATESLDEFDKAYSNNKTYV---E 292
Qy 127 SKDGHFELLGKI---SNVSKAKVYVYNNYSIE-----IKATKYDFHSKTMTPDLY 175
Db 293 SKSEHQKLSELEIKERTSNSDVAKNKLVEIEIDLDSIYNLK-SRYEIQKLITED-- 349
Qy 176 ANINDIYDGLAPAGDMRLFVQNDQKAKIRIEMPE--KIKETYSB---YPVYSSYGN 228
Db 350 KNYNK-----ELSKINKSEKKNIDGLIEWEMKSIKQYKNDADIIISTISQNNN 400
Qy 229 VIELEBGLSKKRPDLTRMES---GKIYDSEKQOVLKDNITLRKGVALKVTYNGP 284
Db 401 EAVI---LKKIEISNESKLSIRKAGGYKS-----LKINEYTKTLSELVKIN-- 448
Qy 285 KTDMLBAGGVYSKEDIKIOKANPULRALSETTIYVADSRNVEDGSTVSLMADGFNI 344
Db 449 --DKISYBQIRERKRSKISLN--RIISB----- 475
Qy 345 IRYGVFTKMDKCEALDKDGNLYTSSKVLFGDKRYTGEDKFNVAIKEDGSMLEFI 404
Db 476 --BKANLELNSKSKLEANKMLNL-----BKQYBQYNN-SVNLQHQHYTKGFV 522
Qy 405 DTKPVNLSMDKNFNPSSNKIYVNPPEFYLRGKISDKGFMELRVNBSYVD---NYLI 461
Db 523 DVKPSSEFVLSEVIRKVEKPEFAV--SISLAALSD-----ITLDNIAKKLIYNLK 573
Qy 462 YGDLIHIDTRDPNI---KLNVKDG----- 482
Db 574 SKNLGRATFLPLNIIKGGKLNISDTRHEKPGIGIASLIDYSTFLPAVYVLAGRTVIV 633
Qy 483 DIMQKMDYRANGPDPKVTMDGNVYLQYGSIDNAKAVGVHVOFLYDNVKEPVNIDPK 542
Db 634 DNMOSALKIATLNSYSFPIVTLTGEV-----VNG 663
Qy 543 GNTS--IEYADGKSVFNINDKNNNGPFGBIQEOHIYINGKEYSFNDIKQIIDTKLNK 600
Db 664 GSLTGGSTYSKQASLT-----GRREIEBELNLENNNSQA---LEQSNKLIENK 710
Qy 601 IYVXDPANNTVKEBILNKDGEVSELKPEHVTYIIONGKEMSTIVSEDEPLVYKGE 660
Db 711 KVVXG-----LDNLCLD-----LTDYI-HGKELKELKIKERLSIDIESEK 750
Qy 661 LEKYQDPGMEISGEBGKQAGVYNLSKQTFPIKVPFKIEBK--KEBENKPTDVSKK 717
Db 751 LKNSYNTSVGEG-----PIKEKIKHLEKLVKEENNA---LKLR 789
Qy 718 KQNPVNHSSOLNESHRE-----EDLOREHSSOKSD-----STKDVATVLDK 759
Db 790 EAN---NNMLIDELERKLDKNSKVLNBERLMSKVKVXASDNLMSSTRIBRYKVM 846
Qy 760 NNISK 765
Db 847 HMEWK 852

```

RESULT 12

US-10-744-616-7

Sequence 7, Application US/10744616
Publication No. US20050026170A1
GENERAL INFORMATION:
APPLICANT: Patti, Joseph M.
APPLICANT: Foster, Timothy J.
APPLICANT: Hook, Magnus A.O.
APPLICANT: Bidinn, Delidre Nl
APPLICANT: Peltine, Samuel L.
TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from *Staphylococcus aureus*
FILE REFERENCE: P06283052/BAS
CURRENT APPLICATION NUMBER: US/10/744,616
PRIOR FILING DATE: 2003-12-24
PRIOR APPLICATION NUMBER: 60/096,815
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: 60/098,427
NUMBER OF SEQ ID NOS: 23
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 7
LENGTH: 1166
TYPE: PRF
ORGANISM: *Staphylococcus aureus*
US-10-744-616-7

Query Match 5.0%; Score 203; DB 17; Length 1166;
Best local similarity 21.6%; Pred. No. 0.0062;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

33 QENESISE--KSFITIDRNISTIRDPENKDKLIKKKFR-----EYDDFTS 78
226 KNPPEKELVANDNDSTHTCPVATPTSVAPKVNAMKRPVAVQPAVAVASNNVDLIK 285
79 ETGRMEYDYKYDDKGNINAYDDGTLEETE-----KLEDEISKYGVLS 125
286 VTKQIK---VGDEKDVNVAADKDIETDTEFTIDNKKKKDTMTINYNKAVIPSDLT 341
126 PSKD-----GHEEILGKISNVSKNAK---VYGNKYKIRIKATKYDFHSK----- 168
342 DKNDPIDITDPSGEVIAGTDPKATKQIYTFDVIDKEDIKSLTLYSYIDKTVPM 401
169 ---TWTF---DLYANIN-DIVDGLAPAGD---NELPVK-DNDQKAEIKIRM-PEKIK 214
402 TSLMTFTAGKETQNTVVDYQDPMVH-GDSNIQSIPLKDLBDRQTEEQIYVNPILKS 460
215 ETKSEY---PYVSSGVNVIHGBGD--LSRNKPNLTM-----ESGKYSNSKQ 261
461 AINIKVDIAGSQVDYGN-IKLGNGSTIIDQNTETIKRVKNSDQQLPQSNRIYDFSYED 519
262 YLTK-DNIIIRKGYALKVTVNPGKTDMLGNGVYSKEDIARIQANPILALSETTIVA 320
520 VTSQPDN---KKSFEENNVAITDFG-----DINSAYI-----IKVSKTTPS 558
321 DSR-NVEDGRSTQSVLMSALDGFNIIRYQVPTFRANDK-----E 359
559 DGEIADIAGTSKRT--DKYGYNAGYSNFIIVSNDTGSGDGYKPEBKLYKIDYVWE 616
360 AIDKGNLVTSQSKVLVPGKDDKEYTGEDKFNVAIIXDGSMLFDITGVNLSMDXNY-F 418
617 DVDKDG-----VQSTDSKERTMANVL-VTLTPPDGT-----TTSVRTDANGHEEF 660
419 NPSKSNKIYV--RNPEFYLRGKIS-----DKGFPNMLAVNESVVDNYLYIGDLHD 468
661 GGLDGEFYTVKPEFTYGLPTKVGITDGEKDSGSSVTVKINCK-----DDMSLD 712
469 -----NTRDFNLIKLVNKG--DINDMGKDYKANGFPDKVTDMOANVYLQGYSL 517
713 TGFYKBPKNAGDYWEDTNDGIDQDANEPIKDVKVT---LDSYGVK-IGTTTTDA 766
518 NAKAVGVYQFLVNDKPEVNIIDPKGNTSIEY--ADGKSVFN--INDKRNAGDFGLQ 572
767 SGK-----YKF-----TDLN-NGNVTVEFTAGTTPYKATTTADKDSNGL----- 807
573 EOHVIYINKEYTSFNDIKQIIDKTINIKI VVKDFPARN--TVKEPI---LNKDNGBVSL 627

808 -----TTGVYIKADNMITLD-----RGFTYTPKXSLGDYVWYNSNDGKQDST 851
628 KPR-VTVIIONGK-BMSSTIVSEDFILPVTKGLEIKGYQDPGMHISGFEKGDAGYVI 865
852 KQIKVTVTLQNEKGEVIGTITDEN-----GK-----YRFD----- 883
686 NLSKOTFIKVPFKIEKKEBENKPFQVSKGGDNQV-----NHSQ--LNSHRKEDLQ 738
884 NDSGKT-KVITEKPAGLTQVTMTED-DKQADGGEVYITTDHDFILNKGYFEDTS 941
739 REHSOKSDSTD 751
942 DSDSDSDSDSDSD 954

RESULT 13
US-10-032-585-7646
Sequence 7646, Application US/10032585
Publication No. US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Roemer D.
APPLICANT: Bo, Jlang
APPLICANT: Charles, Boone
APPLICANT: Howard, Busey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-005-999
CURRENT APPLICATION NUMBER: US/10/032,585
PRIOR FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: Patent In version 3.1
SEQ ID NO 7646
LENGTH: 1881
TYPE: PRF
ORGANISM: *Candida albicans*
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (1881)..(1881)
OTHER INFORMATION: X-any amino acid
US-10-032-585-7646

Query Match 5.0%; Score 203; DB 14; Length 1881;
Best local similarity 20.8%; Pred. No. 0.011;
Matches 172; Conservative 132; Mismatches 300; Indels 222; Gaps 39;

22 KQTTGV---HHQENESIKKSSFTIDN-ISTIDPFRKOLKLIKKKFRVDDF 76
1140 KQATDEIFTKTHITDLOEHAHKQKQSFESERNDIKSNIDANKELSD-NREKLSNLER 1198
77 TSETGRMEYDYKYDDKGNITAY--DQGTDLAEFTEKLEIKSKLYGVLSKDGHPFI 134
1199 KTELANKKATQSEKISDLETVAISDEKSKSLKHIDIKRKKILKLETTIKNBEFTMPFK 1258
135 LKISNVSRNAKVVYGNKYKIRIKATKYDFHSKMTFPLVYININDIVDGLAPAGDRLP 194
1259 KEQLQVNVDRCK-----ELBAC-----LKKLTETKEKINDILIRLEAA----- 1297
135 VQNDQKAEIKIRMPEKIKETKSEHYVSSGVNTEGEGDLSNNKPNLTKMESGKIY 254
1298 KSDHTEKRLSL---LBDTKSB-----SEKNVTKLNE-----QLEK 1334
255 SDSRQOYLLKONIIIRKGYALKVTVNPGKTD-----LBNNGVYSKEDIKQ 305
1335 GERBKE---VRD---IQGQLAKATTDMKIKTTLKVLKKSGLDEKT---NNESVTLKK 1385
306 ANPNLR--ALSETTIVASRNVBDGRSTQSVLMSALDGFNI-----IRQVP- 350
1386 EYENLKEKISILEDDKODITTKYKEIAAQLETKTSNLSSTWLEKTELELKKNVRELTE 1445
351 ---TFRNDKGS---BAIDKGNLVTSQSKVLVPGKDDKEYTGEDKFNVAIIXDGSML 402
1446 ATSELTKQDNNQSLTEIEIKTALTKSSK-----DLEVCAGNKSGLQ-----DSLK 1493

```

Qy 403 FIDTKPVNLSDKXVFNPSKNIYVRNPEFY-----LRGKISDGKFNWELRVNESVVD 457
Db 1494 SVKSLKFNFKNTQOETTSLDELEKQXEVTLQTELRDRISBEKERAMLSRSETV- 1552
Qy 458 NYLIYGDHLINDTDFNKLNVKOGDIDMKGKDYKANGFPDXKVTMDGNNVLTQYSDI, 517
Db 1553 -----IKEXS-----DKISLE-----SKI 1567
Qy 518 NAKAVGHYQFLYNDVKNPEVNIIDPKGNTSIRY-----ADGKSVFNINDKRNNGPDGEI 571
Db 1568 NS-----IKENHSKEITTHNEQKTSILKODIAKLQSODHSAQTOLEDKQKELKA 1618
Qy 572 Q-BOHVINGKEYSF-----NDIKQIID--KTLNKIYVKKPANTVYKSF-----ILKRD 620
Db 1619 SLEKH-----NTSATSIEBKNOIKELSETIKSLKELTSGDALKQSQREYKTLTKNSD 1675
Qy 621 T-----GEVSELEKHRTVTTLONGKMSSTVSEDFILPVYKGLERKGYOPDGEWISGF 675
Db 1676 TESKLERQLEELE--KVMSDLOTDADRKLKGIATERIAL-----KSELET----- 1717
Qy 676 BGKQDAGVNLSDKDTFLKVPYFKIIEKKKEBKNTFPVSKKQNPQVNHQSOLNESHK- 734
Db 1718 --VNSGSLSTSELAALTKTV--KSLERKEBELQPLSGNKSKELEBYIQHSDISEKRAL 1774
Qy 735 EDLQREHSHQSDSTK-----DVTATVLDKNNISKSTTNPNK 773
Db 1775 TDELKEKTKQFPDSDSKKGLTLENDUTST---KKELTEKTYTSRK 1817

```

```

RESULT 14
US-09-742-096-3
; Sequence 3, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DRUHLIE, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-BRYTHOCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773USODIV
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: P. falciparum
US-09-742-096-3

```

```

Query Match 5.0%; Score 200.5; DB 9; Length 1786;
Best Local Similarity 20.5%; Pred. No. 0.015;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

Qy 6 AES--KPKN-LGNGEGSLKQDT-----TGVEHHQENR-----SEIK 40
Db 868 AESVTFPSNILEEIOENTITNDTIEKLEELHENVLSAALNTVOSBEKKEVIDVILEVK 927
Qy 41 EKSEFTI-----DRNISTIRD--FEN-----KOLKGLIKKKFREVDDFTS 78
Db 928 EVAVTTLIEVQAEKESKANTITIEFLELNAVNSNENVAENLEKLEMTYVNTVLDKVE 987
Qy 79 ET-----GRME--EYDYKY-----DDKG--NIAYDGTDLLETYT-----EKLD-- 114
Db 988 ETVISGSLNENNEEDKAFPSHIEPDNVKGIOENLLT---GMFRSIEISIVIOSEBEKVDLN 1044
Qy 115 -EIKSKIYGVLSPSKDGFEILIGKISVSKQAKV-----YGNNGYSIRIKATKQYFH 166
Db 1045 ENNVSSILNDIENKEG--LNLKLNISSTEGVQETVTEHVEQNVYDVDPVPAK----- 1097

```

```

Qy 167 SKMTFPLDYNINIDVGLAAGMRLLPVKNDQKAEIKIRMPKIKETSEPYVSSY 226
Db 1098 -----DQFLGILMERAGL-----XEMFNLIEDVKSSDVLTVEBIDEPQKVEKET 1146
Qy 227 GNVIELEGGD-----LSKRPDNLFTQM-----ESGCIYDSEKQOYLKONILIRKGYALK 277
Db 1147 VSIIEBENIVDVLEBEKEDLTDKMDAVEESITISSDSKERTSINDK---EKDVSLV 1203
Qy 278 VTTNPG-----KTDMLRGNGVSK--EDIAKIQANENL----- 310
Db 1204 VEEVDNDMDBSVEKYLELKNNEBELMDQAVEINDITSKLEBTOELNEVRADLLIKDMK 1263
Qy 311 -----RAISFTIYADSNVEDGR--STQSYLSALDGFNITIRYVFTPKMDKBEADK 363
Db 1264 LKELEKALSE-----DSKXIIIDAKODTLEKYLEEHD-----ITTTLDEVVELKQV 1309
Qy 364 DGNLVTDSKLVLPKQDKKEYTGEDKFNVAIKEDGSMLEFIDPKPVNL-----SM 413
Db 1310 EEDKIBKVSOL-----KQLEBDLKE--VKELKELESELBEDYKELKTIETDILBEKKEI 1362
Qy 414 DKNTFNPSSKNIYVRNDEFTLRGKISDGGFNMLRVNESVVD-----NYLIYGD 465
Db 1363 EKDHFEKEEBEABEIKLEADILKEVS-----SLEVEEKKLEBYHLEKEVEHIIISGDA 1417
Qy 466 HINDTRDNILKLVND--GDIM-----DMGMDYKANGFPDXKVTMDGNNVLTQYSDLN 518
Db 1418 HIKGLEBDLE--EVDLNGSLIDMLKGMELDQDKESLEBDVTTKGERV---ESLKQVL 1473
Qy 519 AKAVGVHYQFLYDN-----VXPEVNIIDPKGNTSIEYADGKSVFNINDK--RN 564
Db 1474 SSALGMDBEQKMTKRRKQAPKLEBVLKBEVYBEPKKIT-----KKYVRFDIKQKRPD 1528
Qy 565 NGPDGEIOBOHIYNGK-----EYTSFNDIKQIIDTLN-----IKIYVKDPRK--N 609
Db 1529 EIVEVEMQDEDIIEBVEEDIEEDIEBDEYKEDIDEDIDIGDKQDEVDLIYQKERRIEK 1588
Qy 610 TTVEKFLNKDTGE--VSELEKPH-----RYVTIONGKMSSTVSEED 651
Db 1589 VKAKKKLEKVEBVSGLKQHVDEVMKYQKIDGVDKESKALSKSDVNVNLKQND 1648
Qy 652 FILPVYKELERKGYOPDGMELISGFEGKQDAGVNL-----SKDTFLKPVK 698
Db 1649 FFSKV--KTFVKKYKFAAPISAVAA--FASVVGFTFSLSSCVTLIASSTYLLSKVDK 1705
Qy 699 KLEBKESKRP-----TQDV-----SKKQNPQVNHQSOLNESHKREDLQR 739
Db 1706 TINKNK--REFYSFVEDIFQNLKHYLQQMKERFSKERNNVIEVT--NKAERKGNVQV 1760
Qy 740 BEHSQSDSTKQVATVLDKNNISKSTTNPNK 773
Db 1761 TNKTEKT-----TKYDKNNKVPKRRTOQSK 1786

```

```

RESULT 15
US-10-415-253-2
; Sequence 2, Application US/10415253
; Publication No. US20040067236A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Joe
; TITLE OF INVENTION: Immunogenic Compositions Comprising
; TITLE OF INVENTION: Liver Stage Malarial Antigens
; FILE REFERENCE: B45250
; CURRENT APPLICATION NUMBER: US/10/415,253
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: PCT/EP01/12349
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: EP00203724.0
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:09:54 ; Search time 181 Seconds

(without alignments)
2186.946 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026
Sequence: 1 KLGRIASKEFKNLGNKRGK.....ATVLDKNNISSKSTNNPNK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	4026	100.0	2119	2	Q9ART5 streptococc
2	4026	100.0	2140	2	Q97RY6 streptococc
3	3793.5	94.2	2144	2	Q94AM8 streptococc
4	3709.5	92.1	2144	2	Q8DOP7 streptococc
5	270.5	6.7	2649	2	Q7RAS7
6	265.5	6.6	1642	2	Q81B84
7	264	6.6	3504	2	Q81L45
8	260	6.5	1850	2	Q7RGP2
9	259	6.4	2757	2	Q7RRR9
10	258.5	6.4	1811	2	Q7REH9
11	257.5	6.4	2661	2	Q7RMS4
12	257	6.4	1389	2	Q7BRP4
13	255	6.3	1777	2	Q813P4
14	253.5	6.3	1127	2	Q9YV76
15	252.5	6.3	2227	2	Q81121
16	252.5	6.3	2723	2	Q7ROB6
17	251.5	6.2	2849	2	Q81HY4
18	251.5	6.2	3381	2	Q812V4
19	251.5	6.2	3519	2	Q81E65
20	250	6.2	1474	2	Q81L02
21	250	6.2	1650	2	Q77328
22	247.5	6.1	5767	2	Q81525
23	246.5	6.1	2033	2	Q81M18
24	245	6.1	2269	2	Q81LA2
25	242.5	6.0	2664	2	Q7RBL0
26	242.5	6.0	3317	2	Q8EMR8
27	241.5	6.0	1033	2	Q81B88
28	241.5	6.0	2694	2	Q7RJ11
29	241.5	6.0	3063	2	Q6IMC1
30	240.5	6.0	2740	2	Q7RFS2
31	240.5	6.0	4433	2	Q81J15

32	240	6.0	1455	2	Q81XG8
33	240	6.0	10061	2	Q81J21
34	239.5	5.9	2569	2	Q81B68
35	239	5.9	2586	2	Q7PDT7
36	238	5.9	1081	2	Q8X1L2
37	238	5.9	3322	2	Q81XLO
38	238	5.9	3628	2	Q968Y0
39	238	5.9	3704	2	Q81XY8
40	237.5	5.9	1104	2	Q7RSO8
41	237.5	5.9	3535	2	Q81C29
42	237	5.9	2511	2	Q81L44
43	237	5.9	5229	2	Q7RTF4
44	236.5	5.9	1114	2	Q97242
45	236.5	5.9	1738	2	Q81A15

ALIGNMENTS

RESULT 1					
Q9ART5	PRELIMINARY;	PRT;	2119	AA.	
AC	Q9ART5:				
DT	01-JUN-2001 (TrEMBLrel. 17, Created)				
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)				
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)				
DE	Serine protease (Fragment).				
GN	Name=prt1;				
OS	Streptococcus pneumoniae.				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_Taxid=1313;				
RM	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=N4;				
RX	MEDLINE=21115976; PubMed=1179332;				
RX	DOI=10.1128/IAI.69.3.1593-1598.2001;				
RA	Witzemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,				
RA	Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,				
RA	Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,				
RA	Langermann S., Johnson S., Koenig S.;				
RT	Use of a whole genome approach to identify vaccine molecules				
RT	affording protection against Streptococcus pneumoniae infection.*;				
RT	Infect. Immun. 69:1593-1598(2001).				
RL	-1 SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by				
CC	an amide bond (by similarity).				
DR	EMBL: AF291699; AKJ9159.1; ..				
DR	HSSP: P00782; 2SRT.				
DR	MEROPS: S08.064; ..				
DR	GO: GO:0009986; C:cell surface; IEA.				
DR	GO: GO:0005618; C:cell wall; IEA.				
DR	GO: GO:0008233; F:peptidase activity; IEA.				
DR	GO: GO:004289; F:subtilase activity; IEA.				
DR	GO: GO:0006508; P:proteolysis and peptidolysis; IEA.				
DR	InterPro: IPR010435; DUF1034.				
DR	InterPro: IPR001899; Gram_pos_anchor.				
DR	InterPro: IPR003137; PA.				
DR	InterPro: IPR002029; Pept_S8_S53.				
DR	InterPro: IPR010259; Prot_inh_S8A.				
DR	InterPro: IPR001680; WD40.				
DR	Pfam: PF06280; DUF1034; 1.				
DR	Pfam: PF00746; Gram_pos_anchor; 1.				
DR	Pfam: PF02225; PA; 1.				
DR	Pfam: PF00082; Peptidase_S8; 1.				
DR	Pfam: PF05922; Subtilisin_N; 1.				
DR	PRINTS: PR00723; SUBTILISIN.				
DR	TIGRfam: TIGR01167; lpxTG anchor; 1.				
DR	PROSITE: PS00847; GRAM_POS_ANCHORING; 1.				
DR	PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN; 1.				
DR	PROSITE: PS00138; SUBTILASE_SER; UNKNOWN; 1.				
DR	PROSITE: PS00678; WD_REPEATS_1; UNKNOWN; 1.				
KW	Cell wall; Peptidoglycan-anchor; Protease.				
FT	NON_TER	1	1		

SQ SEQUENCE 2119 AA; 238227 MW; 51769B7F6B960A6A CRC64;
 Query Match 100.0%; Score 4026; DB 2; Length 2119;
 Best Local Similarity 100.0%; Pred. No. 2e-161;
 Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIASEKFNKLGNGKSGSLKCDITGVGHHQENBESIKESKSFITIDNISTIRDPENK 60
 DB 1313 KLGEIASEKFNKLGNGKSGSLKCDITGVGHHQENBESIKESKSFITIDNISTIRDPENK 1372
 QY 61 DLKLLIKKKFRVNDPTSETGKMEHYDYKDDKNIIAYDGTDLLEYTEKLDRIKSK 120
 DB 1373 DLKLLIKKKFRVNDPTSETGKMEHYDYKDDKNIIAYDGTDLLEYTEKLDRIKSK 1432
 QY 121 YGVLSPSKDGHPHLLGKISNVSNNKAVYGNKYKSEIKATKDDPFASKMTFIDLYANIND 180
 DB 1433 YGVLSPSKDGHPHLLGKISNVSNNKAVYGNKYKSEIKATKDDPFASKMTFIDLYANIND 1492
 QY 181 IVDGLAFAGDMRLPVKNDQKABIKIRMPKIKETKSEYPVSSYGNVIELGEGDLSKN 240
 DB 1493 IVDGLAFAGDMRLPVKNDQKABIKIRMPKIKETKSEYPVSSYGNVIELGEGDLSKN 1552
 QY 241 KPDNLTNKGSGKIYDSKQOYLKDNIIIRKGYALKTTTPRGKTDMLKNGVYSKEDI 300
 DB 1553 KPDNLTNKGSGKIYDSKQOYLKDNIIIRKGYALKTTTPRGKTDMLKNGVYSKEDI 1612
 QY 301 AKIQANPRLALSETTIYADSRNVEDGRSTOSVMSALDGNFIIRYQVTFPMNDKGBA 360
 DB 1613 AKIQANPRLALSETTIYADSRNVEDGRSTOSVMSALDGNFIIRYQVTFPMNDKGBA 1672
 QY 361 IDKQGNLTVDSSKVLFGKDKYETGDEKFNVEAIKEDSMLFIDTPVNLSDMXNYPN 420
 DB 1673 IDKQGNLTVDSSKVLFGKDKYETGDEKFNVEAIKEDSMLFIDTPVNLSDMXNYPN 1732
 QY 421 SKSNKIYVRNBEFYLRGKISDGKGFNWRARVNESVVDNLTLYGDLIHIDTRDPNKLANK 480
 DB 1733 SKSNKIYVRNBEFYLRGKISDGKGFNWRARVNESVVDNLTLYGDLIHIDTRDPNKLANK 1792
 QY 481 DGDIMDMGKQYKANGFPDKVTDMDGNVLTQYGSDLNAKAVGHYQFLYDNVKEBVND 540
 DB 1793 DGDIMDMGKQYKANGFPDKVTDMDGNVLTQYGSDLNAKAVGHYQFLYDNVKEBVND 1852
 QY 541 PKGNISIRYADGKSVYFNINDKRNNGFDGEIOEONIIYNGKETSFNDIKQIIDTLNLIK 600
 DB 1853 PKGNISIRYADGKSVYFNINDKRNNGFDGEIOEONIIYNGKETSFNDIKQIIDTLNLIK 1912
 QY 601 IVKQDPANNTYKCEIILNKDNGESELKPHRYVTITLONKEMSTIVSEEDFILPVYKGE 660
 DB 1913 IVKQDPANNTYKCEIILNKDNGESELKPHRYVTITLONKEMSTIVSEEDFILPVYKGE 1972
 QY 1973 LEKGYQFDGWEISGEGKKDAGVINTLSKDTFIKPVFKIIEKKEBENKPTDVSKKCN 2032
 DB 2032 LEKGYQFDGWEISGEGKKDAGVINTLSKDTFIKPVFKIIEKKEBENKPTDVSKKCN 2032
 QY 721 POWNHQSINSHRKKDLOREHSQKSDSTKQVTAIVLDKNISSTSTNNPK 773
 DB 2033 POWNHQSINSHRKKDLOREHSQKSDSTKQVTAIVLDKNISSTSTNNPK 2085

RESULT 2
 Q97RY6 PRELIMINARY; PRT; 2140 AA.
 AC Q97RY6;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Serine protease, subtilase family.
 GN OrderedNames=SP0641;
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
 RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Hatt D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,
 RA Holtapple E.R., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Peidilyum T.V., Angiolini S.V., Dickinson T.,
 RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae."
 RL Science 293:498-506(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 an amide bond (by similarity).
 DB EMBL, AE007373; AKT74791.1; -.
 DB PIR, P95074; P95074.
 DB HSSP, P00782; 2SBR.
 DB MEROPS, S08.064; -.
 DB TIGR, SP0641; -.
 DR GO, GO:0009386; C:cell surface; IEA.
 DR GO, GO:0005618; C:cell wall; IEA.
 DR GO, GO:0008233; F:peptidase activity; IEA.
 DR GO, GO:0004289; F:subtilase activity; IEA.
 DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR010435; DUF1034.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR003137; PA.
 DR InterPro: IPR000209; Pept_S6_S53.
 DR InterPro: IPR010259; Prot_inh_S8A.
 DR Pfam, PF06280; DUF1034; 1.
 DR Pfam, PF00746; Gram_pos_anchor; 1.
 DR Pfam, PF02225; Pa_1.
 DR Pfam, PF00082; Peptidase_S6; 1.
 DR Pfam, PF05922; Subtilisin_N; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRFAMs; TIGR01167; LPTYG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
 KW Cell wall; Complete proteome; peptidoglycan-anchor; Protease.
 SQ SEQUENCE 2140 AA; 240426 MW; PA44AD8E2938B314 CRC64;

Query Match 100.0%; Score 4026; DB 2; Length 2140;
 Best Local Similarity 100.0%; Pred. No. 2e-161;
 Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIASEKFNKLGNGKSGSLKCDITGVGHHQENBESIKESKSFITIDNISTIRDPENK 60
 DB 1313 KLGEIASEKFNKLGNGKSGSLKCDITGVGHHQENBESIKESKSFITIDNISTIRDPENK 1393
 QY 61 DLKLLIKKKFRVNDPTSETGKMEHYDYKDDKNIIAYDGTDLLEYTEKLDRIKSK 120
 DB 1394 DLKLLIKKKFRVNDPTSETGKMEHYDYKDDKNIIAYDGTDLLEYTEKLDRIKSK 1453
 QY 121 YGVLSPSKDGHPHLLGKISNVSNNKAVYGNKYKSEIKATKDDPFASKMTFIDLYANIND 180
 DB 1454 YGVLSPSKDGHPHLLGKISNVSNNKAVYGNKYKSEIKATKDDPFASKMTFIDLYANIND 1513
 QY 181 IVDGLAFAGDMRLPVKNDQKABIKIRMPKIKETKSEYPVSSYGNVIELGEGDLSKN 240
 DB 1514 IVDGLAFAGDMRLPVKNDQKABIKIRMPKIKETKSEYPVSSYGNVIELGEGDLSKN 1573
 QY 241 KPDNLTNKGSGKIYDSKQOYLKDNIIIRKGYALKTTTPRGKTDMLKNGVYSKEDI 300
 DB 1574 KPDNLTNKGSGKIYDSKQOYLKDNIIIRKGYALKTTTPRGKTDMLKNGVYSKEDI 1633
 QY 301 AKIQANPRLALSETTIYADSRNVEDGRSTOSVMSALDGNFIIRYQVTFPMNDKGBA 360
 DB 1634 AKIQANPRLALSETTIYADSRNVEDGRSTOSVMSALDGNFIIRYQVTFPMNDKGBA 1693

```

QY 361 IDKQGNLVTDSKVLFGKDDKEYTGEDKFNVAIKEDGSMLEFIDTRFVNLSMDKQVFN 420
DB 1694 IDKQGNLVTDSKVLFGKDDKEYTGEDKFNVAIKEDGSMLEFIDTRFVNLSMDKQVFN 1753
QY 421 SKSNKLIYRNPEFLRGKISDKGFGNWEIARVNESVVDNYLYIGDLHIDNTRFNIKLVN 480
DB 1754 SKSNKLIYRNPEFLRGKISDKGFGNWEIARVNESVVDNYLYIGDLHIDNTRFNIKLVN 1813
QY 481 DGDIMDMGKDYKANGFPDXTDMDGNVYLQIGYSDIARAKAVGVHYQFLYDNVKEPVNID 540
DB 1814 DGDIMDMGKDYKANGFPDXTDMDGNVYLQIGYSDIARAKAVGVHYQFLYDNVKEPVNID 1873
QY 541 PKGNSTIRYADGKSVVFNINDKRNNGPDGEIOBOHIYINGKRYTSFNDIKOIIDTLNIN 600
DB 1874 PKGNSTIRYADGKSVVFNINDKRNNGPDGEIOBOHIYINGKRYTSFNDIKOIIDTLNIN 1933
QY 601 IVVADPARNTVKEEILNKDGEVSELKPHRYTVIIONGKEMSTIVSEEDFILPVYKGE 660
DB 1934 IVVADPARNTVKEEILNKDGEVSELKPHRYTVIIONGKEMSTIVSEEDFILPVYKGE 1993
QY 661 LEKGYQPDGWEISGEGKKDAGYVNLSDPTFKVFKIEKKERENKPTFVSKKON 720
DB 1994 LEKGYQPDGWEISGEGKKDAGYVNLSDPTFKVFKIEKKERENKPTFVSKKON 2053
QY 721 PAVNHSOLNESHKREDLQREBHSOKSDSTKDYATATVLDKNISSTKSTTNPNK 773
DB 2054 PAVNHSOLNESHKREDLQREBHSOKSDSTKDYATATVLDKNISSTKSTTNPNK 2106

```

RESULT 3

```

09S4M8 ID 09S4M8 PRELIMINARY; PRT: 2144 AA.
AC 09S4M8
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2004 (TREMBlrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase precursor Prta.
GN Name=Prta;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCB1_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3.B;
RX MEDLINE=21585565; PubMed=11728722;
RA Behe G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,
  Zyk G.;
RT "The cell wall-associated serine protease Prta: a highly conserved
  virulence factor of Streptococcus pneumoniae.";
RL FEMS Microbiol. Lett. 205:99-104(2001).
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
  an amide bond (by similarity).
DB EMBL; AF127143; AAD48399.1; -.
DR HSSP; P00782; 2SPT.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001317; PA_721.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA_1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.

```

```

DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LpxTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASIN_HIS; UNKNOWN; 1.
DR PROSITE; PS00138; SUBTILASIN_SER; UNKNOWN; 1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN; 1.
DR Cell wall; Peptidoglycan-anchor; signal.
FT SIGNAL 1
FT CHAIN 20 2144
FT PRTA.
SQ SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;
Query Match 94.2%; Score 3793.5; DB 2; Length 2144;
Best Local Similarity 94.7%; Pred. No. 1.3e-151;
Matches 732; Conservative 16; Mismatches 24; Indels 1; Gaps 1;
QY 1 KKGAIASKEPKNINAGKESGLKQDTTGVEHHQENBESIKESSTFIDNISTINDPNK 60
DB 1339 KKGAIPESEKFNILKAVKQDSILKETAFAVENNLTAVNQSIGKSLFNIHRTISTINDPNK 1398
QY 61 DKKLIIKKKRFVNDFTSEPTGKEMEDYKDYKDDKNITIAVDGTDLEETEKLDEIKSI 120
DB 1399 DKKLIIKKKRYQEDFPV-GGRTYERDRTYDDKGNITIAVDGTDLEETEKLDEIKSI 1457
QY 121 YGVLSPSKDGHEILGKISNYSKNAKYVYGNYSKISIKATKYDFHSKTMTEFDLYANIND 180
DB 1458 YGVLSPSKDGHEILGKISNYSKNAKYVYGNYSKISIKATKYDFHSKTMTEFDLYANIND 1517
QY 181 IYDGLAFAGDMRLFYKNDQKKAELIKIMPEKIKETKSEYFVYSYGVNIEYAGEGLSKN 240
DB 1518 IYDGLAFAGDMRLFYKNDQKKAELIKIMPEKIKETKSEYFVYSYGVNIEYAGEGLSKN 1577
QY 241 KPDNLTKNESGKIYDSEKQOYTLKDNITILRGYALKVTYTPGKTDMLRGKGVYSKED 300
DB 1578 KPDNLTKNESGKIYDSEKQOYTLKDNITILRGYALKVTYTPGKTDMLRGKGVYSKED 1637
QY 301 AKIQANFNLRALSETTIVADSRNVEDGRTQSVLMSALDGNIIIRYQVFFRKMDKGEA 360
DB 1638 AKIQANFNLRALSETTIVADSRNVEDGRTQSVLMSALDGNIIIRYQVFFRKMDKGEA 1697
QY 361 IDKQGNLVTDSKVLFGKDDKEYTGEDKFNVAIKEDGSMLEFIDTRFVNLSMDKQVFN 420
DB 1698 IDKQGNLVTDSKVLFGKDDKEYTGEDKFNVAIKEDGSMLEFIDTRFVNLSMDKQVFN 1757
QY 421 SKSNKLIYRNPEFLRGKISDKGFGNWEIARVNESVVDNYLYIGDLHIDNTRFNIKLVN 480
DB 1758 SKSNKLIYRNPEFLRGKISDKGFGNWEIARVNESVVDNYLYIGDLHIDNTRFNIKLVN 1817
QY 481 DGDIMDMGKDYKANGFPDXTDMDGNVYLQIGYSDIARAKAVGVHYQFLYDNVKEPVNID 540
DB 1818 DGDIMDMGKDYKANGFPDXTDMDGNVYLQIGYSDIARAKAVGVHYQFLYDNVKEPVNID 1877
QY 541 PKGNSTIRYADGKSVVFNINDKRNNGPDGEIOBOHIYINGKRYTSFNDIKOIIDTLNIN 600
DB 1878 PKGNSTIRYADGKSVVFNINDKRNNGPDGEIOBOHIYINGKRYTSFNDIKOIIDTLNIN 1937
QY 601 IVVADPARNTVKEEILNKDGEVSELKPHRYTVIIONGKEMSTIVSEEDFILPVYKGE 660
DB 1938 IVVADPARNTVKEEILNKDGEVSELKPHRYTVIIONGKEMSTIVSEEDFILPVYKGE 1997
QY 661 LEKGYQPDGWEISGEGKKDAGYVNLSDPTFKVFKIEKKERENKPTFVSKKON 720
DB 1998 LEKGYQPDGWEISGEGKKDAGYVNLSDPTFKVFKIEKKERENKPTFVSKKON 2057
QY 721 PAVNHSOLNESHKREDLQREBHSOKSDSTKDYATATVLDKNISSTKSTTNPNK 773
DB 2058 PAVNHSOLNESHKREDLQREBHSOKSDSTKDYATATVLDKNISSTKSTTNPNK 2110

```

RESULT 4

```

ID 08DOP7 PRELIMINARY; PRT: 2144 AA.
AC 08DOP7;

```


DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
 GN Name=prta; OrderedLocustNames=spr0561;
 OS Streptococcus pneumoniae (strain ATCC BAA-295 / R6).
 CC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus
 CC NCBI_Taxid=171101;
 OK (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=12429245; PubMed=11544234;
 RX DOI=10.1128/JB.163.19.5709-5717.2001;
 RA Hockley B.S., Alborn W.B. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA Dehoff B.S., Battem S.T., Fritz L., Pu D.-J., Puller W., Geringer C.,
 RA Gilmour R., Gabe J.S., Khoja H., Kraft A.R., Lagace R.B.,
 RA Leblanc D.J., Lee L.N., Letkowitz B.J., Lu J., Matsushima P.,
 RA McHenry S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
 RA Norris P.H., O'Garra M., Peery R.B., Robertson G.T., Rokey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R. Jr., Skatrud P.L.,
 RA Glas J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
 RL J. Bacteriol. 183:5709-5717(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 an amide bond (By similarity).
 CC EMBL: AB008434; AK9365.1; -.
 DR PIR: A97942; A97942.
 DR HSRP: P00782; ZSRT.
 DR MEROPS: S08.064; -.
 DR GO: GO:0009986; C:cell surface; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0004289; F:subtilase activity; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR010435; DUF1034.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR003137; PA.
 DR InterPro: IPR000209; Pept_S8_S53.
 DR InterPro: IPR010259; Prot_inh_S8A.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF06280; DUF1034; 1.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR Pfam: PF02225; PA; 1.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR Pfam: PF05922; Subtilisin_N; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR TIGRPFAM: TIGR0167; LPRYG_anchor; 1.
 DR PROSITE: PS0847; GRAM_POS_ANCHORING; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN; 1.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN; 1.
 KM Cell wall, Complete proteome; Peptidoglycan-anchor.
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC03A0C CRC64;

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
 Best Local Similarity 92.4%; Pred. No. 4,5e-146;
 Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 1 KLGRIAEKFKNGKXGSLKQDTGVHHQHEESIKKSSFTIDRNSTIDFENK 60
 DB KLGRIAEKFKNGKXGSLKQDTGVHHQHEESIKKSSFTIDRNSTIDFENK 1397
 QY 61 DLKGLIKKKFRVDDFTSETGRMEBYDYKYDDKGNIIAYDGTDLLEYETKLDIKSKI 120
 DB DLKGLIKKKFRVDDFTSETGRMEBYDYKYDDKGNIIAYDGTDLLEYETKLDIKSKI 1457
 QY 121 YGVUSPSDGHPELIGKISNYSKAKYVYANNYSIEIKATKYDPHSTKTMFDLYANIND 180
 DB YGVUSPSDGHPELIGKISNYSKAKYVYANNYSIEIKATKYDPHSTKTMFDLYANIND 1517
 QY 181 IVDGLAFAGDRLFLVYKDNDDKAKIKIMPEKIKETKSEYFVYSSYGVNIEIGSDLSKN 240
 DB IVDGLAFAGDRLFLVYKDNDDKAKIKIMPEKIKETKSEYFVYSSYGVNIEIGSDLSKN 1577

QY 241 KPDNLTMSSEKTVSDSEKQVYLKDNITLRGVALKTVTPRGKTMLENGYISKEDI 300
 DB KPDNLTMSSEKTVSDSEKQVYLKDNITLRGVALKTVTPRGKTMLENGYISKEDI 1637
 QY 301 AKIQANPLRALSETTYIADSRNVEDGRSTOSVMSALDGFNIIIRYQVTFEAMDKGBA 360
 DB AKIQANPLRALSETTYIADSRNVEDGRSTOSVMSALDGFNIIIRYQVTFEAMDKGBA 1697
 QY 361 IDKGNLVTDSSKLVLPQGDKEYTGEDKFNVEAIKEDGSMFPIOTKPVNLSMDKNYFNP 420
 DB IDKGNLVTDSSKLVLPQGDKEYTGEDKFNVEAIKEDGSMFPIOTKPVNLSMDKNYFNP 1757
 QY 421 SSKNTIYVNPPEFYLRGKISDKGFMNRLRVNESVVDNLYIGDILINDTRDFNKLANK 480
 DB SSKNTIYVNPPEFYLRGKISDKGFMNRLRVNESVVDNLYIGDILINDTRDFNKLANK 1817
 QY 481 DGDINDWGMKDKFKANGFPDVKVTDMQGNVYLQTVSYDLNAKAVGVHQPFLYDNVKEVNID 540
 DB DGDINDWGMKDKFKANGFPDVKVTDMQGNVYLQTVSYDLNAKAVGVHQPFLYDNVKEVNID 1877
 QY 541 PKGNTSIEYADGKSVYVFNINDKRNNGPDGEIODEHIYINGKRYTFPNDIKOIIDTLNIX 600
 DB PKGNTSIEYADGKSVYVFNINDKRNNGPDGEIODEHIYINGKRYTFPNDIKOIIDTLNIX 1937
 QY 601 IYVVDPAFNTVYKEFTLNKQDGEVSELKPHRYVTYITONGKEMSTIYSEDFILPYVYKGB 660
 DB IYVVDPAFNTVYKEFTLNKQDGEVSELKPHRYVTYITONGKEMSTIYSEDFILPYVYKGB 1997
 QY 661 LEKGQDFDGMELSGEGKKDAGVYINLSKDTFIKVPFKIEBKKEBNKPTFVSKKON 720
 DB LEKGQDFDGMELSGEGKKDAGVYINLSKDTFIKVPFKIEBKKEBNKPTFVSKKON 2057
 QY 721 PQVNSQINESHRKEDLOREHSQKSDSTQVATVILKNNISKSTTNPNK 773
 DB PQVNSQINESHRKEDLOREHSQKSDSTQVATVILKNNISKSTTNPNK 2110

RESULT 5
 ID Q7RAS7 PRELIMINARY; PRT; 2649 AA.
 AC Q7RAS7;
 DT 01-MAR-2004 (TREMblrel. 26, Created)
 DT 01-MAR-2004 (TREMblrel. 26, Last sequence update)
 DE 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN Name=PY06422;
 OS Plasmodium yoelii yoelii.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCBI_Taxid=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=17XNL;
 RX PubMed=12368865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angiloul S.V., Suh B.B., Koell T.W., Pierce M.,
 RA Silva J.C., Ermolaeva M.D., Allen J.B., Selengut J.D., Xoo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Bidwell S.L.,
 RA Shallow S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabdi A., Cummings L.M.,
 RA Florens L., Yates F.R. III, Raine J.D., Sinden K.E., Harris M.A.,
 RA Cummings D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Garnci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RL Nature 419:512-519(2002).
 CC -1- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 CC EMBL: AAB01002172; EAA18637.1; -.
 DR InterPro: IPR011591; Botulinum.
 DR ProDom: PD001963; Botulinum; 2.

KW Hypothetical protein.
 SQ SEQUENCE 2649 AA; 309810 MW; E207F344643AC24 CRC64;

Query Match 6.7%; Score 270.5; DB 2; Length 2649;
 Best Local Similarity 20.6%; Pred. No. 0.0052;
 Matches 206; Conservative 160; Mismatches 327; Indels 307; Gaps 49;

1 KLGEIAEKPKVKNLGN-----GKSGSLKDDTGVGHHHGHQENBESIKESGPTIDRISTID 56
 1044 KNNLIERNVRDNDNLHGVNKAKKNNKIDSNILYNKSGVDGVYQVRASYITTFETFTIIS 1103

57 FE-----NKDKKLIKKKPREVDFTSETGKMEBYDYKYDDKGNIIAYD----- 102
 1104 FEDKIKILQNERLKKIKLEKLYDEKRYKNDY-KMEKIKKQNDY--NLFEATDGHIEK 1159

103 -GTDLEETKDELKSKIVGVLSPSKQHFELIGLISVNSKRAKYVGN--NYGSIHRIK 159
 1160 LHCELENNKSKONELIKINL-----KDKMKKILIELSQICNN-NVNSHEHLNHTMLDVK 1211

160 ATKVPFHSKWTFFDLVYANINDIVDGLAFAGDM-----RLVFDKNDQKAE-IKIRMPKRI 213
 1212 NSENNYKHEHTNNNDVHSLNDSITDCKKIKKQIVLQDQLPALKNEIKQMEFLSKVVELN 1271

214 KETKSEYVYSSYGAVIELGSGDLSKPKPDNL----- 245
 1272 KLLNKKKNNIMEYERSIDKLEKSIDKLEKQNLKLEQICDLKERNIMLEKAAQIRDESS 1331

246 -----TQMSG-----KLYSD----- 256
 1332 NTTSSIDSGTTINNEIKIMEKIEALYKDKIKLKSNELEKTKNINILTLTKTSNEGS 1391

257 ---SEKQOYLKDNILIRKGVALKYTYNPKGIDMLBNGVYSKEDIQKAPMLRAL 313
 1392 IELNKKIKLCKENKGLQKNYKESI-----NDLKANIKYD-EDILKENSIIKPSII- 1442

314 SETTIYADSRVDEGSTQSVLMSALD--GNNIIR--YQVTFKRNKDG-----EALD 362
 1443 --TFEDLNNEKIEHNTVQKIDESTQIDDSYDMARKKGGYDISHKONSSECYQDIKTEID 1500

363 KD-----GNL-----VTDSKLVLPCKDKKEYT--GDKFENYALIKEDSGMLFIDTKPV- 409
 1501 KMMTEKCKTPNKLKETOIDDITNLVLSVCKKEMTEKNQDDYN--IIEEDNLKINRNSNF 1557

410 NLSMDKXNPNPSKSNKIYVRNPEFY-----LRKLSIDKGG-----FWMEILRVNESVVD 457
 1558 NNRKENIYTNLLKNNKCYVNVNIFIDISIRANLQWMPFNANGNSFNMYEVNNE--H 1614

458 NYL-----YGDHLIDNTRDPNFK--LAVKQGDIMMGMDYKANGPPDKYTTMDGAV 508
 1615 NYTNMKNKINAYSNSHLSN--NISSEHVSQNTESIQ-----NYNKKYVNTKYPADIBV 1664

509 YLQGTGYSDLNKAVGVH-----YQFLYDNYKPEVNIID 541
 1665 YNRN-----NLQFISLHNDENNRRTISNGTISNVSERKDNSKTYVNNNSKEITIEEN 1719

542 KGNTSIEYADGKSVFENIDK-----RNNGPDGEIQEOHIYINGKEYTSFNDIKQIIDK 595
 1720 NNNNSIKNSD-----NVNKKYECCLKINBESANNESTLNNTEENNSTDLKNIIYE 1773

596 TLNFKI---VVDPAKNTTVKGFILN-KOTGVSLEKPRVYVTLONGEMASTIVS-EE 650
 1774 DNNLIAVNNKIIENY-NDQDLKNYILNSQCTNNMNEHKPND-NMTVEKKKKGEMWIIDIN 1831

651 DFIIPVYVGGELKGVQFDGMBISGPEGKAGVILNSKOTFLKPVFKKIKBEKCEBENP 710
 1832 NEVLPYTK--IEKCVLSD-----EKBACTKKNNRK-----KSCQKNSKNSKT 1872

711 -----TPDVSK-----KQDNPOVN-H-----SQLESER 733
 1873 VNNMNRRTYNIYKRESEESIKNKTEFPANCKNPILMERTKNGKLDLVLNDSIKLVKNKI 1932

734 KEDLOREHSGOKSDSTKQVT---ATVLDKNNISKSTTN 770

DB 1933 KEELKGNIS-----KQITNLEITITIKKKNPISKSLN 1966

RESULT 6
 ID 081B84 PRELIMINARY; PRT; 1642 AA.
 AC 081B84;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, last annotation update)
 DE Hypothetical protein MAL8P1.29.
 GN Name=Mal8P1.29;
 OS Plasmodium falciparum (isolate 3D7).
 OC Burkarya; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Seeger K., Murphy L., Hattie D., Berriman M., Pain A., Hall N.,
 RA Quail M., Barrell B.,
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL844507; CAD51123.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1642 AA; 194753 MW; 701F5D2BB2BE8097 CRC64;

Query Match 6.6%; Score 265.5; DB 2; Length 1642;
 Best Local Similarity 20.4%; Pred. No. 0.0048;
 Matches 190; Conservative 133; Mismatches 313; Indels 297; Gaps 46;

4 ELAESKPKVLGNGKESLKKDDTGVGHHHGHQENBESI--KESGPTIDRISTIDPEENK 60
 104 EINRNNRNNRNNRNNNSNSPIMDBKEKTYNNRTIMLKHKKRYDPDYNTS-----DEK 158

61 DLKLIKKKKPREVDFTSETGKRM-----EYDYKDDKGNIIAYDGTDLLEYETKDLDE 115
 159 NIKR-----KKINDINIKKKKKMILPSYBENDKSNKGSFISD-----DE 201

116 IASKIYGVLSPSKQHFELIGLISVNSKRAKYVGNVYSIEIKATKYDFHSKWTFFDLY 175
 202 M-----DNKSNVYNSIGK-----NY-STENNNDKNNFHISRL--EKM 235

176 ANINDVGLAFAGDMRLVYKNDQKKAIRKMPKIKETSEYVYSSYGAVIELGSG 235
 236 DKINN-----EKKEY-----STKSDIDLRVVISHSRGY-----NNNIDKLLKH 275

236 DLKKNPDNLTKMESKRIYDSEKQOYLKDNILIRKGVY----- 275
 276 NLIARNKSNLVNPDNRK-----RKENPFIHANYMKCKENKNTNINSYDIENHDEKDK 330

276 -----LKYTTNPGK---TMYLGN-----GYSKG-DIAKQKAPNLRALS 314
 331 NEILANNLIDINSYBERKLSYNNKRENDPFKIGISKOPNVDSKERYKYEKDHNIYNDLS 390

315 ETTIYADSRNVEGR-STQSVLMSALDGNFIIRYQVTFPRQNDK----- 357
 391 DCKSGIDRTNVYDNNSSSHVSLSNL--FNESNPFMLDEKKIRDKYKGVKLIKESLERNS 448

358 -----GEAIDXD--GNLYTD-----SSTLVLPFGDKKEYTGBDKFNVNAIKS 397
 449 ELYENELKEGETHKOBLKVLVDMTIIKOINEKKKVVILYHKDYS-----INERSISS 503

398 DQSGMLFIDTPKPVNLSMDKXNPNPSKSNKIYVRNPEFYLAGKISDKGSGFMELAVNESVVD 457
 504 EN-----IDIGPTNRSFLVGNNTYKDMTHINENNDR-----DKNNND--INN--D 544

458 NYLIVGDLHDNTRDPNFIILNAYKGDIMMGMDYKANGPPDKYTTMDGAV-----YL 510
 545 NYIINDNININNNYVNVK--ESNEBIF-----YANGLSKEDINIVQNIIRIKKIKKS 596

511 QGTGYSDLNKAVGVHGYQFLYDNY-KPEVNIIDPKGN--TSIEYADGKSVFENIDKRN 565
 597 NSGNVNIIDAN-----LSDGISNDEANILKAVTNSMDYTKLYDDDK-----GENTK 642

566 GPDGEIQEOHIYINGKEYTSFNDIKQIDKTNLIKIVVDPAKNTTVKPIILKQDGBVS 625

```

Db      643 TF-----KSAVDXKXKLLIENDTLMHNNNN-----QKEMIND--EKTLMGINS PK 691
Qy      626 ELKPRVATYTIQ---XGKEMSTIVSEDPILPVYKGLKGYQPDGWEISGEGKCD- 680
Db      692 DLDRKNSNIDIOVNNNNNNNNNDTTFYDAMEYSIGNRKEOYNKVOED-HVIMGVESKNN 750
Qy      681 -----AGVINTLSKDTPL-----KPVF-- 697
Db      751 MGTNCSNNKQYINNDRNHLKEXYFDMMKPGVVDNIINIKENELVLOKNEKKOIFNK 810
Qy      698 ---KXIEKKEEENK-PTFDVSKKQDNPQVNH-----SQLNESH--KEDLQRE 740
Db      811 SHGNEKLEBPLKOPNNKQVAVTNILKQKVVNDIDHEDVSVNIGRIDENRKKKKENILNS 870
Qy      741 EHSQKSDSTKD-VTATVLDKNNISSKSTNNPN 772
Db      871 INQEKDIGKNNIINKTSYSKNNILSKSYAKEPH 903

```

RESULT 7

```

ID      081145 PRELIMINARY; PRT; 3504 AA.
AC      081145,
DT      01-MAR-2003 (TREMBLrel. 23, Created)
DT      01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Hypothetical protein.
GN      ORFName=PF14_0404;
OS      Plasmodium falciparum (isolate 3D7).
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxId=36329;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=2255705; PubMed=12368864; DOI=10.1038/nature01097;
RA      Gardner M.J., Hall N., Pung B., White O., Bertram M., Hyman R.W.,
RA      Carlton J.M., Patil A., Nelson K.B., Bowman S., Paulsen I.T., James K.,
RA      Bisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyte S.,
RA      Chan M.S., Nene V., Shallow S.D., Sub B., Peterson J., Angluoli S.,
RA      Perea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA      Martin D.M., Fairlamb A.H., Fraunholz M.J., Roop D.S., Ralph S.A.,
RA      McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA      Venter J.C., Carucci D.D., Hoffman S.L., Newbold C., Davis R.W.,
RA      Fraser C.M., Barrell B.
RT      "Genome sequence of the human malaria parasite Plasmodium
RT      falciparum."
RL      Nature 419:498-511 (2002).
DR      EMBL; AB014822; AAN37017.1; -.
DR      HSP; 0931C2; IMW.
KW      Hypothetical protein.
SQ      SEQUENCE 3504 AA; 408308 MW; B8454D48D55B84F0 CRC64;

```

```

Query Match      6.6%; Score 264; DB 2; Length 3504;
Best Local Similarity 21.2%; Pred. No. 0.014;
Matches 196; Conservative 131; Mismatches 346; Indels 252; Gaps 43;
Qy      11 KNLGKSGSLKCDTGVVHHQEN--EESIKER-----SFTIDRNISTIRPENKDL 62
Db      1432 KNSHNSSEDMITEBKGNKSPQENIDMIIPKNGENKNSQONSQSHNIMTEKNGKNS 1491
Qy      63 KKLKKRREVDPTSETGRMEEDYKTDK---GAILADGDTLLEYTEKLDKETS 118
Db      1492 QO-----NSDHNIMTEEYKKNKSOENTDNTIBGYGNKNSKKKEEDIAS 1539
Qy      119 KLYGV-----LSPSK-GHPBILGKISVSKNAKVYV-----NNYSIRIKATKVP--- 165
Db      1540 ---YELDKRISHENQEHPTYESRINKGFHNVVDNIYVSGNEKGISVKNISBIIID 1597
Qy      166 -HSKMTPTDLYANINDIYDGLAPADQMLFTVQNDQKAEIKIRMEKIKETKS-YP 221
Db      1598 GKGNITODDILEEN-----GKNKFEHIEBEDKISDXTQ-KSKISHAKBGFPTP 1647
Qy      222 YVS-----SYGNVIELGEGDLKNNKPDNLTQMESGITY-----SDSEKQGYAL 264

```

```

Db      1648 YESGKXKISDENVEVENSIDINTNKOQ--EVESKQIPFNQNNIKHSSDNKKLNPK 1705
Qy      265 KNNIILRKGYALKVTTYNPKTD-KLEGNGYVSKEDIQKANKPULR-----ALSETTI 318
Db      1706 QNNILINESODKNI DVHN--KLDKILKNEVTSDESLKTKKENGTRKNGKGINNEK 1763
Qy      319 YADSRVY-----EDGSRQSVLMSALDFNIRYOVTFPKNN----- 355
Db      1764 EEEKENVADDETIGGKEBNTESDCLKIQKISNENKNNILYTDNTNKKSYAAGGTHG 1823
Qy      356 -----DKGBAID--KQGNLV----- 368
Db      1824 ENDETNGTNISNDGDKRVKTIQOYISKGNITLOENKBDIIPSVTINNSIGDKVXENLS 1883
Qy      369 -TDSSTLVFGHDKKRYTGDEKNNV-----EAKEGSNLFDITK-----PVYLS- 412
Db      1884 PEDIKKEVAAKNNIOMITSEDELGTQKONERKEKXSPNGVEHNEQENDKIIGEVNLN 1943
Qy      413 MDKRYFNPSSKNTIYVRNPEFYLRK-ISDKGPFNBLR--VRESVVDNY-----L 460
Db      1944 KNNNESNIGNSDTI---NQHLMGKNIHKKGVNSETMENTNSGTNNISNEQEPKNI 2000
Qy      461 IYGDLIHIDTROPNITLAVKDGIDMDGKKDYKANGPPKVTMDGNNVYLQTSYDLNAX 520
Db      2001 IRGD-DIKOMNENVXKLEDETGN-----NIKINYNDNAKILNELIKNGQTSODAD 2052
Qy      521 AVGVHQLYDVK-PVVNIDPKGNTSEYADKSVVFNINDRNNGPDEIOBHIYIN 579
Db      2053 DISTNSDKRQDQKNNENNIHKKNNVNE---KDKI--SNDKEDNIVPE-HKEDIIIS 2105
Qy      580 GKRYTSFNDIKQIIDTFLNIKIVVKDPARTTYKRFILNKDTESEVELKPHRYVTYIONG 639
Db      2106 DNKKKEFDNVLEIPKIGNI-----LDDKRTTRQVEEKSIGQDSMENNVSST--NDG 2157
Qy      640 KEMSTIVSEBDFILPVYVYKGLKGYQPDGWEISGEGKQAGVYNLSK----- 689
Db      2158 KQIH---IQEBDI-----KENIINNVDNRHSESKNNIHIPEPNYVEKEIKK 2204
Qy      690 ---DFTIKVVPKRIEKEEENKPT-----FOVSKKQDNPQVNHQSLNESHREKDLQRE 741
Db      2205 EIAHDHDKKFKELIQGDSNKNSEFNSMENILVDNAODK-----NISKLTNDLHDQ 2256
Qy      742 HSQKSDSTQDVTATVLDKNNISSKS 766
Db      2257 KGTNDSVVE-----HNSDKT 2273

```

RESULT 8

```

ID      07RG2 PRELIMINARY; PRT; 1850 AA.
AC      07RG2;
DT      01-MAR-2004 (TREMBLrel. 26, Created)
DT      01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DE      01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE      Hypothetical protein.
GN      Name=PY04304;
OS      Plasmodium yoelii yoelii.
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxId=73239;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=17XNL;
RA      Carlson J.M., Angluoli S.V., Sub B.B., Kooij T.W., Perea M.,
RA      Silva J.C., Ermolaeva M.D., Allen J.B., Selengut J.D., Koo H.L.,
RA      Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Bidwell T.V.,
RA      Shallow S.O., van Aken S.B., Riedmiller S.B., Feldblum T.V.,
RA      Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.A.,
RA      Florens L., Yates P.R. III, Raine J.D., Sinden R.B., Harris M.A.,
RA      Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA      van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA      Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,

```



```

QY 186 AFAGDMLPYKNDQ-----KKAETIKRMBPKI-----KETSRYPVSSYGNTEL 232
DB 773 -----TAKENDQNLVYKSKISEIKNHYNDOISINTYEEKKONYDOSKRYTITSI 824
QY 233 GEGDLSKN-----KPDNLTMB-----SGCIYDSEKQOYTLKDNITILKGYALKV 278
DB 825 KQDSTKTIINAVNMDQFLSKYDKYNFENCNCENVDSEHTQOTELTNKIKAEVSDKL 884
QY 279 TTY-----NPGKTMLEANGVYSRE--DIACIQKAPNLRALST-----TIYAD 321
DB 885 SIYENKPNKSKSILNKTNNBIKRYKONINTLKVDYKICESTYKBSIKNPHNKTILKE 944
QY 322 ---SRNVEDGRSTQVLSALDGFNIIIRYQVFTPDNDKGRABIDQNL---VTDSSLYL 376
DB 945 KANQNKITIKETNISIKSYDKPE---NLINKINELDKAF--KQASLNDYESNNNELMQ 999
QY 377 PGHDKREYTGDEKFNVAIKEDSMLPID-----TKPVLSMDK--NYNPKSKN--KIYVRN 430
DB 1000 YFNANLKANLGKKNEN-----MLYHQDEKEKAVNNIIQKIEDINQILPIEIAIYT 1050
QY 431 PEYLAGKISDKGFWNEL-----RVNRSYVDNYLIYGDIAIDTRDPRIKILNYQGD 483
DB 1051 STYNISEELNBIQKIBSLNTVLEKVTNTNPNFKIKELKLYNPSDFQ---KEGN 1105
QY 484 IMDWAKMDYKANGFPDKVTMDGNVYLOQYSDLNAXAVGVHYOFLYDNPBRYNIDPK- 542
DB 1106 I-----KTNENVKIKODIKNVNQIDHINLELEIKKSENYVDEKTAQINLEKY 1157
QY 543 -----GNTSIEVADGK--SVAPNINDRN-----NGFDGSIQEQHIYINGEYTFNDIK 590
DB 1158 ADTAISNDVNEGIEKQONIVTKDKKKNIYBEINKLSISKIE---KQTSLEKVK 1212
QY 591 QI-----IDTILN 598
DB 1213 GINLSYQNLGTYFLBQIDEBKKSEHTITQAMEYMEDLNKKKSOBIENENGICENDIN 1272
QY 599 IKIVNDPAMNTVKEBILNKDTGE--VSELKPRVVTYIQNGKMSSTIYSEDFILPVY 657
DB 1273 KQMEVNTISHDDKQVHTTSIKNNENISDIRNKSILKI-ID-----PSRSDINDI- 1322
QY 658 KGELEAGY---QPDGHEISGFBGKQAGYVITNLSKDTFIPVFKKIEKTEBENKPTFDV 714
DB 1323 KQELQNVSSSQKNSIBINQYLSK--IANYINILKNIKIKQINKVKEAYNBIK- 1375
QY 715 SKKQDPQVNHQSOLNESHKEDLOREHESQKSDT---KQVTAATVLD---KONNISKST 767
DB 1376 NKNQINDELSNEXLIKIKYKODISLKECKSKIESITLDDKQIDQCIQIKVLTHTLSEET 1435
QY 768 TNN 770
DB 1436 NIN 1438

```

```

RA Shallom S.J., van Aken S.B., Riedmiller S.B., Feldhym T.V.,
RA Cho J.K., Quackenhush J., Sedegah M., Shoaib A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.B., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Frazer C.M., Hoffman S.L., Gardner M.J.,
RA Garucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DB EMBL: AABL01001591; BAA17051.1; -.
KB Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1811 AA; 210756 MW; E76A90F7392BCA0A CRC64;

Query Match 6.4%; Score 258.5; DB 2; Length 1811;
Best Local Similarity 21.7%; Pred. No. 0.011;
Matches 179; Conservative 117; Mismatches 312; Indels 215; Gaps 40;

QY 11 KNLNGKESGLKQDTTGVHHRHOENBS-----IKKSSFTIDRNISTIRDFENKD 61
DB 555 KNINSGHGSNIHVNRRIRIHNNNDQNTVSNHILKQDETQKIDINDVQNTATYGF 614
QY 62 LKTLIKKKPREVDFTSETGRKEHBYKYDDKGNIAIYDGTDLER-----ETKDE- 115
DB 615 ILS-DKEDFKLLFTPLNDKGSFPLFEKMLKDSKQWIDILQOREYINNNILEKKEE 673
QY 116 -----IKSKIYGVLPSEKQGFPELIGKISVSKN-AKVYVGNANKSIEIKATKTFPH 166
DB 674 KKYVLEKCIINNKIKELMCGRVKGLMNV--KLCNMEKKIINAFSLNIIKNNIHVTLIDY 731
QY 167 SKTYTF-DLYANINDIYDGLAFAGDKRLPYKNDQKAEIKIMPEKIKETKSYPIVSS 225
DB 732 RSAMCYODIYA-----QLYVYGLNSK---KIEMTKKINEFK----- 764
QY 226 YGNVIELGEGDLSKNRPDLTKMESGK-----IYSDSEKQOYTLKONITLKGAL- 276
DB 765 --NNID-GVANSQENSDAKKQDSSKKKGYVSHLSYFNKSKH--KQKYSIN 815
QY 277 ---KVTYNPQKTMLEGVNSKEDIAKIQKANPMLRALSETTIVASRVNVEDGRSTOS 333
DB 816 NDTKQTT-EPSSKD-----NLAKKAEHBRITANNNEDEGRKLEN 853
QY 334 VMSALDGFYITIKYQVTFK-VNDK-----GEALIDQGLVTPSSKIVLFGKD--KQYT 385
DB 854 KQKG-----NINRTVNTKRLNEQYQGRNSDETDKSNIRTDKANNPNFGKNSYNEKT 908
QY 386 GEDKFNVEAIKEDGSMFLIDTKPVNLSMDKNVFNPS-KSNKIYVRNPEF---YLKGLIS 440
DB 909 VEDK-----EIKETKNSVNYVYISNNSYENNSYLDNNSFKNKLNKSEIV 953
QY 441 DKGFNWEIKRVNRSVDN--YLIYGDLIHNTDFENIKLVNKGDIIMDKGKQYKANGP 498
DB 954 EKSNNSN---IKKKKVDNCGSGSFGSP-----FMFK-----SDVNRKG-- 987
QY 499 DXYTMDGNVYLOQYSDLNAXAVGVHYOFLYDNPBRYNIDPKGNTSIEYADSKSVFN 558
DB 988 SQIESNNNSAIIVPKSDDSNB--GSDYSKYNNKEDKROMKKNSSKKSNNMAFKK-IPN 1043
QY 559 INDKRNNGPGEIIO---EQHIYINGKEYTSFNDIKOITLTKINIKIVKDPARTVYKEF 615
DB 1044 LSRKKKKKNGGSDSPSLKSKDKIKKQDNTMDSVANNEDK-----YNNESSNQ 1091
QY 616 ILNKDTGEVSELKPRHRYVTYIQNG-----KMSSTIVSEEDTILPVYKGELEAGYFDG 669
DB 1092 SSNEEDGHINNYKN-----NGNKKKANKNSNGNLNKSLSLEFQKDFDEEFGPI- 1142
QY 670 WEISGFEKKDAGVYNLSQDTFIPVFKKIEKKEBENKPTDVSKKQDPQVNHQSOLN 729
DB 1143 -----LKQEVDPHLIHDVDEKSIIFERFDEDR-----FDIGTKNSNSCSNS--N 1187

```

Qy 730 ESHKEDLQREHRSQSDSTKDVAT-VLDKNNISKSTNNP 771
Db 1188 DSNSSD-----SKSIRSDQDTQTQTKSSSKSEKQ 1224

RESULT 11

07RMS4 PRELIMINARY; PRT: 2661 AA.

AC 07RMS4
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Rhodoty protein (Fragment).
GN Name=PY02104;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed1236865; DOI=10.1038/nature01099;

RA Carlton J.M., Angitoli S.V., Suh B.B., Kooli T.W., Pertea M.,
RA Silva J.C., Brmolaeva M.D., Allen J.B., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.R., Riedmiller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabhi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.B., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

CC EMBL; AAB01000575; EAA21526.1; -
DR InterPro; IPR011561; Pox 11 rel.
DR InterPro; IPR006499; ReticulocyteBP.
DR ProDom; PD014111; Pox_11_rel; 1.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
FT NON TER 2661 2661
SQ SEQUENCE 2661 AA; 312269 MW; 0BD02AD0B71253B CRC64;

Query Match 6.4%; Score 257.5; DB 2; Length 2661;
Best Local Similarity 20.0%; Pred. No. 0.018;
Matches 179; Conservative 151; Mismatches 294; Indels 271; Gaps 41;

Qy 22 KQDTTGEHHQENESIK-EKSPFTIDNIS-----TRDPENK-- 60
Db 427 KKLLETTTKANKENBEYVLEKEIRELFKYSDEVAEKGYBELKALKETIKQIYNKE 486
Qy 61 -----DLKLIKKKFREVDDFTSYGKMEBYDYKYD-----KGNIIAYD 101
Db 487 YIKKAIIDLKAIENNNKYIDELGKNTPOIREYVKKOTIVSTIKSELISIKYIVELY 546
Qy 102 DGTDLBETRLDITKSK-----LYGULSPKDGHPRELIGKISVSKY-- 144
Db 547 NEUSSVQENTTIDPIKKNKELETLSKSEIDVNNYKIKNNKIEBVELHNLNITSKNELSNTL 606
Qy 145 -----AKVYVG--NNYKSIEIKATKYDFHSTKTFDLYANINDIYDGLAFAGDMRLFYVGN 198
Db 607 ILLEIKIIFGEIDN-----DLNKITAKEFNKQ--ELSKIND-----YTKEN 647
Qy 199 DQ-----KKAELIKIRPEKI-----KETYSRYPVYSYGVNIELGEGDLK-- 239
Db 648 DQLSVYOSKILIRKHYNDQINVNTKGBEAKVNDOSKRYAKKIPIRGEGTSKFXINEVT 707
Qy 240 -NKPNDNLTKR-----SGKLYSPSEKQVYLLKNNI-----LRGYLAKYTTNPG 284
Db 708 AMKNELLKJLKKYIDFGATYKAKANSHEKITEVKKIRIISKYLVANNYERK--FNNS 764

Qy 285 KIDMLKGNVYSKE--DIKIQKANPNBALSETTYADSRVNGCRSTQVLSALD-G 341
Db 765 KKLITETDKSIEBEKYNINWTLKKVDYIKVCYH-----NELSNFRNKQTILKDKLDNR 819
Qy 342 FNIRY-----OVFEK-----NNDKGAIDKQNNLY-----TDSKVLVF-----G 378
Db 820 INTVKENNSIDLTYNTKFNILITDKTELDKPTTVALADHESNNNELKYPNNLKANLG 879
Qy 379 KQDK-----RYTGBDKFNVFAIYEDSGMLFIDTPVNLSDMKYFNPFSKSKLYVRRPEFY 434
Db 880 KKKKNNMLYQOFGSKKKAIDIKKQADINENYSKIEVALYASIVNISDEMERIESIES 939
Qy 435 LKSGKISDKGFWMLRVNNSVVDNLTLYGDLHIDNRDNITKLVNDSQGLIMKGMQDYKA 494
Db 940 LNTQVYER-----VKTNTVNLMEIKKGLKQYDHDGKENIK----- 977
Qy 495 NGEPDPTVMDGNVYLQTSYSLNKAAGVHYOFLDYNKPEVNIDPKGNTSIEYADGS 554
Db 978 --YDPEINKIKNEI-----KTVG-----QIDQHNKLEIKKSGSKNDELKIGT 1021
Qy 555 VVF-NINDKRNGPGEIOEHYINGKEYSPN-DIKQIIDKTNLKIVKDPANNTV 612
Db 1022 DEFENVYDK-----TYNKPKEIEEKIKNI-VTKIDKKNIYI 1059
Qy 613 KEFLINKQGEVSEIKPHRYT-----VTIQNGKMSSTIVEEDPILFVYGELEKGYQ 666
Db 1060 K---INKLIDETSEIKKNTSLKKYKVDINVSYGSGIKLFLQID----- 1101
Qy 667 PDGWEISGFBGKKDAGVYINLSKDTFIKPV--FKKLEKKEBKNEKPTD-----VSKKKON 720
Db 1102 -----EKKKKGHHIK-SMEERYMDLNIKKKSGQIEBEMKINMDIKYINKEMEA 1151
Qy 721 POWNSQLNESIRKEDLQREHRSQSDSTKDVATVATVLDKN--ISSKSTNNPN 772
Db 1152 LNIISDDKQYH--NISKKBEKISD-----IHNSIKTIQEFSTESNIN 1194

RESULT 12

07RPS4 PRELIMINARY; PRT: 1389 AA.

AC 07RPS4
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY01465;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed1236865; DOI=10.1038/nature01099;

RA Carlton J.M., Angitoli S.V., Suh B.B., Kooli T.W., Pertea M.,
RA Silva J.C., Brmolaeva M.D., Allen J.B., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.R., Riedmiller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabhi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.B., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

DR EMBL; AAB01000391; EAA20892.1; -
DR InterPro; IPR011591; Botulinum.
DR ProDom; PD001963; Botulinum; 1.

KW Hypothetical protein.
SQ SEQUENCE 1369 AA; 164816 MW; 21B51CA630F5D9F CRC64;

Query Match 6.4%; Score 257; DB 2; Length 1369;

Best Local Similarity 20.4%; Pred. No. 0.0089;
Matches 193; Conservative 139; Mismatches 288; Indels 326; Gaps 48;

```

QY 17 KEGSLKQDTGVVEHHQ-----NEBSIKKSSFTIDRNSTIND 56
DB 354 KKEAFQKRNKSSSEHLEKNIPIYKESILKRSKDTQTSYFKESIKKOKSSNNYNNKGVTS 413
QY 57 FENNOLKELIKKRFREVDFTSETGRMEEDYKYDDKGNIIAYDDGTLEYTEKLEDEI 116
DB 414 EIKNGKQIKKNCNKHVFSHSDT-----EKVEY-----SNIS-----DLVNGKNDX 457
QY 117 KSKTYGVLSPSKQDFELIGKISNYSKAKATYGA-NYSIETKAT-----KY 163
DB 458 KLTNYGL---SKEDY---SNDKNNQDTREYVYDNEBKTDIDIOVNDKRVLDISPLYTHKY 510
QY 164 DFHSK-TWTPDLVANIINDVGLAPAGMRLFYKNDOKKAEIKR---MPEKIKETSE 219
DB 511 KYRBNIIITIKIKIKIEKI-----EINLNTITVQKIPGRILBNVILPHK----- 556
QY 220 YPVYSSYGNVI-----ELGEGDLSKKNKPDNLTKESGKIYSDSEK-----QOYL 263
DB 557 -----YENDLIKQIKTISEENENDIKTKKE---TKMEATSYQQSDKNSYQPSMDNV 607
QY 264 LKDNIIILKGYALK-----TTYPNGKIDMLBGNVYSKED- 299
DB 608 IKKKNIDIKTSIKIYVILPLANNINHEIINERLLINSTYB-----MYEKENTL 657
QY 300 -----IAKIQKANPMLRALSETTIYASRNVEDSGSTQVL-MSALDGFN---IIRYQ--V 349
DB 658 LPTDTYILQ---SKSISIKKEMMYISIKKCVNGIYLPILANNVLDNFKNMIMISQNNN 715
QY 350 FTFK-----MADKBAIDKO-GNLYTDSKLVLFKQDKKAYSGDKFNVAIKEDGML 402
DB 716 STFKPYVSLVYNDK---IDREIGNLI-VTNKLI-----NKIKIKKLPKRGYTI 759
QY 403 -----FIDPKPYNLSMDKN-----YENPSKNTKY 427
DB 760 INDICPRFPDKNNVLCVNNRKSNSGNIQIGMLAIDHIGVETANYIIFMASNFTI 819
QY 428 VRNP-----EFTLRGKISDGKGFNMLRVNBSVDNLYLYGDLHIDN 469
DB 820 VKSSLDLPFPHIFLBOVVYFDPIBKMLLAKFIMHPIAITBAICINLCFL----- 872
QY 470 TRDNIIKLANVODGIMKGMQDYKANGPDDKVTMDGNVLYLOTGSDLNAAVGVHYOFL 529
DB 873 ---FNINAQI---DSLFLY---KRLKAVQIKENDQNNPFTFLYTSYB-RVQNNNGHCEYF 923
QY 530 YDNVSP-----EVNIDPKGNTSIRYAD--GKSVFENINDKKNNGPDGEIOBOHIYI 578
DB 924 LNVSKPLFRPOSERELN-----NISHADVEGHTNLCNTSDEIRIKCEBIIKCKKEIE 976
QY 579 NGKETSFNIDIKQIIDLTLNIRIVKVPARNITVVEPILANQDTGVSBLKPHRVTVIQN 638
DB 977 NG-----IK-LIKKTNDNDPFIYDARQSI--FLVN-----AIKNDALIKNIEN 1018
QY 639 GKEMSSITVSEEDF-----ILPVYKGELEK-----GYOF-----DGM 670
DB 1019 NYDKKKQIFYTHSTYKQFODYYINILNRVNOQLQKQKHEDYSRRREVETLQSNIRPQ 1078
QY 671 EISGFEKGKAG-YVINLSKDTFIKVFKEKLEKKEB-----ENKPTFVSKCKQNPVN 724
DB 1079 DVKSYFGKNDNNITSQGLNQ-----FDKCKEKRBBGIIKSSNNKMGQNDKNTDNYQNS 1131
QY 725 HSQANESRKEDLQREHSQKSDSTKDVATVATLAKNNISSKSTNN 770
DB 1132 KQTESDKKPEQNIYTDENIQMLN-----NNFSNTNQVNN 1165

```

RESULT 13

```

O813P4
ID O813P4 PRELIMINARY; PRT; 1777 AA.
AC O813P4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE Hypothetical protein PF01095w.
GN Name=PF01095w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Alkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornby T., Holroyd S., Horrocks P.,
RA Humphrey S., Jagers K., James K.D., Johnson D., Kethnath A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moulé S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultana J.B., Craig A., Newbold C., Barrall B.G.;
RT Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.;
RL Nature 419:527-531(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrall B.;
RA Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL929353; CAD51583.1;
DR InterPro; IPR011591; Bactulium.
DR Prodom; PD001963; Bactulium; 1.
KW Hypothetical protein.
SQ SEQUENCE 1777 AA; 213320 MW; 244467CFPI90C522 CRC64;

Query Match 6.3%; Score 255; DB 2; Length 1777;
Best Local Similarity 20.7%; Pred. No. 0.015;
Matches 186; Conservative 152; Mismatches 324; Indels 236; Gaps 44;

QY 4 EIAESKPNLNGKSGSLKQDTGVVEHHQ-----ESIKKSSFTIDRNSTINDP 57
DB 977 ELDDDKKKKL--DEENELDDKK--KGLDEENELDDKKQGLDEENELDDKKKL--DE 1030
QY 58 ENK--DLKLI-----KKKRFREVDFTSETGRMEEDYKYDDKGNIIAYDDGT 104
DB 1031 ENELDDDKKGLDEENELDDKKQGLDEENELDEKRRKKQMBDNLIDDKKGEIYVDHNT 1090
QY 105 DLEYTEKLEDEIKSI---YGVLSPS-KDGHF-----EILKISNVSQNAK----- 146
DB 1091 FLOTENHLLNEKTKIQDYNIIEALKEKMSKILLSETKIKIENDENDIKRQSOIEN 1150
QY 147 VYTGNNYSIEIKATKYDPHSTWTFEDLYANIINDVLDGLAPAGMRLFYKNDOKKAEIK 206
DB 1151 IYRRNSMMIDNNISYKSNVYTKTFPNDNNE----- 1184
QY 207 IEMPEKIKETSEYVSSYGNVIELGEGDLSKKNPDLTKESGKIYSDSEKQOYLMD 266
DB 1185 ---EDKIQKNTKITYTDISM-----LTNNKSSSIYNSKEDIIINQEIYSNKG 1234
QY 267 NI-----ILKGYALKVTTYNPDKTDLBGNVYSK-----EDIAKIQK-ANPN 309
DB 1235 NTFNNDIKEDLTVNVEKNNNFNNISYKQCSNIVYINSNKFLMDYTESIIKIQNDINEK 1294
QY 310 LRALS--FTTYIADRNVEDSGSTQSLM--SALDGFNIIRYOVTFKMDKGEALDXGN 366
DB 1295 VKLAKQREIPIYEEKGNFEKRNELHLAKENVLNNQNIIKOR--ENELAKKEEELKQKKE 1352

```

```

OY 367 LVTSSKLYLF-----GKODKEY-----TGEDKENYBAIKEDGSMLEPIOTKPYNLGM 413
DB 1333 ILLSEKENTINYSLSNCKINKDKLEYNKLEADVTGKGDINMSYKFSKSHL--SDTN--NIOP 1409
OY 414 DKQYFNPSKSNKIYYRNPEFYLRGKISDGKGFNMBLAV-----NESVDNYLLY-G 463
DB 1410 FENDIN-----NDKLYEDNISSTY--GNNLSHDNNKMYDNNLSGDNMSHDNNMADNNI CHEN 1465
OY 464 DLHINTEDFNKILAV-----KQGDINDGMKDYRANGFPDXYTMDQGVY----- 509
DB 1466 NISYNNPSYNTMKLSBELBENNENKKNIDY-----YDKMFVDPNNININHNFNISKI 1521
OY 510 ---LQGYSDLNAKAVGHYGFLYDVKPEVNI DPKGNSTIEYADSKAVENINDRRNG 566
DB 1522 SNDISISGMNDMB-----LHNDLV-----NNSLKEIETMYKLSRDESEINT 1562
OY 567 PDGEIOBQ-----HIYINGK-----EY--TSFNDIKOIIDKTIN 598
DB 1563 LKLKLEKQRTBEKMKNNINQNLNDISDMANNIYINGNISGVYOTSMHVLBELPEKFLN 1622
OY 599 IKIIVKDFARNTVGEFLINDTG---EYSELKPHRYVYTIQNGKMSSTIYSBEDFT-- 653
DB 1623 LS-EVED---NEKMSIISNELVSLYKELNNIK-----BEYNNIVLKKQEFIGN 1667
OY 654 -LPVYKGELEKGYQFQDWEISGPEGKQAGVYINLSKOTPIKVPFKCIEKKKEBEKPTF 712
DB 1668 LLINFLANDLNNRYL---KANYEKEGVHKSQIILISREDFIKELQNTLNEKQLKE----- 1719
OY 713 DVSKKQDPQVNHSQLMBSHR---KEDLOREBSHOKSDSTQVATVATLDDKNNISKSST 767
DB 1720 -ISYKQMLKRN--QINDTYKLKNRKSLSLTVELKQ--DIKFLNEBVLKCKEAVTLST 1772

RESULT 14
OY 09YVT6 PRELIMINARY; PRT: 1127 AA.
OY 09YVT6
AC 09YVT6;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, last annotation update)
DB Hypothetical protein MSV156.
GN Name=MSV156;
OS Melanoplus sanguinipes entomopoxvirus (MeSPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OC NCB1_Taxid=83191;
RX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99102612; PubMed=9847359;
RA Afonso C.L., Tulman B.R., Lu Z., Oma B., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus."
RL J. Virol. 73:533-552(1999).
DR EMBL: AF063866; AAC97677.1; -.
DR PIR: T28317; T28317.
KW Hypothetical protein.
SQ SEQUENCE 1127 AA; 134265 MW; F18SDALD5A3FE7D1 CRC64;

Query Match 6 3%; Score 253.5; DB 2; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.0098;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;

```

```

OY 162 KYDFHSKMTFPLXANINDIVDGLAAGMELFVK--DNDOKKA-BIKIRMEPKIKET-- 216
DB 368 LFDNDIOQLKLANDITTEONNKITD--FPNNSTRIPKEKLDTEYKLDIDIKNNNLQKLEBSYK 425
OY 217 ---KSEY---PVSSGVAVIELEGGDLKKNRPDNLTKESGKIYDSEKQOYLKONII 269
DB 426 KIDQETREYKKNKINNEYNDIIEKNNNLQKLEBENCKIDQETREYKKNKINKEV--NDIIE 483
OY 270 LKGYALKVTTNPG-----KTDMLKNGVYSKEDIKIQKANPNIRALSETTYAD 321
DB 484 LKNNNLQKLEBENKININDLTKLNDIEBNTBLEFNLTANISDFDKSRERIAKLN--TEYEQ 541
OY 322 SRR--VEDRSQSVL-----MSALDGNRIIRYQVFTFRKNDKGAIDKXGLVYDSSK 373
DB 542 LKDLQLENTKNTLEKLSDNKLSLE-----QLYDSKNTL--GDIDLYNSLTKCN- 592
OY 374 LVLFGKODKRYTGBDFNVEAIKED---GSMLEPIDKPYNLSDMKYFN---PSKSNKIY 427
DB 593 ---DKIDYFSNIEKFDIYVLENKFIQNLDSINKIINDQPKREYINSKIDSKNELS 648
OY 428 VANPEFY-----LRKISDKGCFNMBELRVESVVDNYLYGDLHINTRD 472
DB 649 TWFDFIFNAKQJASITNNIENISNKIXD-----INFEIISNEDSKEL-LDEIRK 698
OY 473 FNILKLVKQDIDMDGMKDYRANGFPDXYTMDQGVY--YLQGYSDLNAKA-----V 522
DB 699 YKQQPD-KIKDMANREVSFE--NTLQKQIDSKSNINELTNADITNTQANDLDDKLANV 756
OY 523 GVHYOFLYDNYVP-----EVNIDPKGNSTIEYAD--GKSVAVFINDDRRNGPDGEIOEHI 576
DB 757 GSBFKNLYNNASDLDDITQKNNDEKYLQNTLEKKNQSIENIDVNN----- 805
OY 577 YINGKYSFPNDIKOIIDKTINIKIVKDFARNTVGEFLINDTG---EYSELKPHRYV 633
DB 806 FI--KELIKFNTE--TNKSLN-ELLTND---DINDKIKLYKELNKISTNNILKIKYK-- 855
OY 634 VTIQNGKMSSTIYSBEDFTLPVYKGELEKGYQFQDWEISGPEGKQAGVYINLSKOTF- 692
DB 856 NEIDVNEKLSVIERLOFINFLSIEFQOQ-----STSH-----INFLANTLA 900
OY 693 -IKVPFKCIEKKKEBEKNTPTPVSKKQDPQVNHSQLMBSHRKEDLOREBSHOKSDSTQ 751
DB 901 GINDVLANKLNTKIMADTTERGDTNIR-----DEIKNQISENICKQXFNKNE--KD 950
OY 752 VTAIVY-----LDKNNISKSSTTN 770
DB 951 LAKGLISFNDKLNKYNISAGYTEYN 974

RESULT 15
OY 081121 PRELIMINARY; PRT: 2227 AA.
OY 081121
AC 081121;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DB Hypothetical protein.
GN ORName=PF11_0354;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCB1_Taxid=36329;
RX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22255705; PubMed=12368664; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung B., White O., Bertzman M., Hyman R.W.,
RA Carlson J.M., Pahn A., Nelson K.B., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyte S.,
RA Chan M.S., Nene V., Shallow S.J., Sub B., Peterson J., Angiulo S.,
RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:11:30 ; Search time 44 Seconds

(without alignments)
1690.354 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGIABSKFKNLGNGKSGS.....ATVLDKNNISSKSTNNPK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2140	2 P95074	serine proteinase,
2	3709.5	92.1	2144	2 A97942	metalloproteinase,
3	253.5	6.3	1127	2 T28317	ORF MSV156 hypothe
4	250	5.8	3724	2 T18444	hypothetical prote
5	235.5	5.8	3724	2 T18427	hypothetical prote
6	231.5	5.8	2269	2 T28677	rhodopsin protein -
7	230	5.7	2401	2 T18440	hypothetical prote
8	228	5.7	4550	2 T18429	hypothetical prote
9	221.5	5.5	1711	2 A24594	probable major sur
10	219	5.4	1640	2 SAZOK1	major serine prote
11	218	5.4	1631	2 S05603	lipoprotein (impor
12	218	5.4	1546	2 G90603	reticulocyte-bind
13	217.5	5.4	1252	2 B42771	DNA-directed RNA p
14	215.5	5.4	1252	2 A45597	membrane nucleas
15	214	5.3	1125	2 E90598	hypothetical prote
16	212.5	5.3	2166	2 G70163	serine/threonine-s
17	212.5	5.3	2485	1 H71621	protein with 5'-3'
18	212	5.3	1188	2 A16866	microbial collagen
19	211.5	5.3	1104	1 A36866	RBSA-H3 antigen pr
20	209	5.2	1558	2 B32885	hypothetical prote
21	207.5	5.2	1441	2 B89824	hypothetical prote
22	207	5.1	1191	2 B97116	chromosome segrega
23	207	5.1	1191	2 C71622	hypothetical prote
24	206.5	5.1	1622	2 AB1717	probable cell surf
25	206	5.1	1526	2 A45605	matr-parasite-in
26	204.5	5.1	1516	2 E71619	RAD2 endonuclease
27	204.5	5.1	2500	2 G71609	hypothetical prote

30	204	5.1	1245	2 D71613	GAF domain protein
31	203.5	5.1	3394	2 T18501	hypothetical prote
32	203	5.0	1166	2 T28680	fibrinogen-binding
33	203	5.0	4981	2 T18489	hypothetical prote
34	202.5	5.0	763	2 G97026	superfamily 1 DNA
35	202.5	5.0	1939	2 T18372	repeat organellar
36	201	5.0	1712	2 C71618	hypothetical prote
37	199	4.9	888	2 E71608	ATP-dept. acyl-CoA
38	199	4.9	980	2 E71606	hypothetical prote
39	197	4.9	1386	2 AC1533	surface protein (l
40	196.5	4.9	2523	2 T18477	hypothetical prote
41	196	4.9	1714	2 E71609	Ser/Thr protein K1
42	195.5	4.9	2380	2 T18504	hypothetical prote
43	195	4.8	1008	2 T18508	hypothetical prote
44	195	4.8	1302	1 JG6009	surface-located me
45	195	4.8	1365	2 T30822	Imp1 protein - Myc

ALIGNMENTS

RESULT 1

P95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIG
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #ext_change 09-Jul-2004
C/Accession: P95074
R/Retellin, H.; Nelson, K.E.; Paulsen, I.T.; Bisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holczapf
nson, T.; Hickey, E.K.; Holt, I.B.
Science 293, 498-506, 2001
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venier, J.C.; Dougherty, B.A.; Morris
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; MUID:21357209; PMID:11463916
A/Accession: P95074
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-2140 <KUN>
A/Cross-references: UNIPROT:Q97RY6; GB:AE005672; PDB:AAK74791.1; PID:G14972117; GSPD
A/Experimental source: strain TIGR4
C/Genetics:
A/Genes: SP0641

Query Match	100.0%	Score 4026	DB 2	Length 2140
Best Local Similarity	100.0%	Pred. No. 1.9e-164		
Matches 773	Conservative	0	Mismatches 0	Indels 0
			Gaps 0	
QY	1	KLGIABSKFKNLGNGKSGSLKDDTGVHHOENESIKKSSPTIDRNISTIRDPENK 60		
DB	1334	KLGIABSKFKNLGNGKSGSLKDDTGVHHOENESIKKSSPTIDRNISTIRDPENK 1393		
QY	61	DLKKLKKKFRVDDPTSETGKRMERYDYKYDDKNTIAYDDGTDELEYETKLDKLSKI 120		
DB	1394	DLKKLKKKFRVDDPTSETGKRMERYDYKYDDKNTIAYDDGTDELEYETKLDKLSKI 1453		
QY	121	YGVLSKSGKHPEIILKISNVSKNAKVVYNNYKSIIEIKATYVDFHSKTTDPYANIND 180		
DB	1454	YGVLSKSGKHPEIILKISNVSKNAKVVYNNYKSIIEIKATYVDFHSKTTDPYANIND 1513		
QY	181	YVDGLAPADMDLFPYQNDNQKAEIKIYRPEIKETKSYPPVSSYGVNTEIGDGLSN 240		
DB	1514	YVDGLAPADMDLFPYQNDNQKAEIKIYRPEIKETKSYPPVSSYGVNTEIGDGLSN 1573		
QY	241	KPDNLTQKSGKIVSDSEKQYLLKONIILKGYALKVTTYPNGKTDMLGNGVSKEDI 300		
DB	1574	KPDNLTQKSGKIVSDSEKQYLLKONIILKGYALKVTTYPNGKTDMLGNGVSKEDI 1633		
QY	301	AKIQANPVLRAISSTTITADSRNVEDGRTSVLSALDGNITIKYVTFPMNDKGA 360		
DB	1634	AKIQANPVLRAISSTTITADSRNVEDGRTSVLSALDGNITIKYVTFPMNDKGA 1693		
QY	361	IKDQGLVYDSSGLVFGKDDKEYGDEFNVAIKEDSGMLPDTKPVNLSDKKYFNP 420		

Db 1694 IDKQGLVTDSSSLVLFQKDKKEYTGDBKFNVAIKEDGSMLEPDTKPVNLSMDKXFNFP 1753
Qy 421 SSKNKTYYNPEFYLGKISDKCGFWMELRVNESVVDNVLIIYGDHIDNTRDNRNKLNVK 480
Db 1754 SSKNKTYYNPEFYLGKISDKCGFWMELRVNESVVDNVLIIYGDHIDNTRDNRNKLNVK 1813
Qy 481 DDGIMQMKDVKYANGPPDKVTDMDGNVYLQTSYSDLANAKAVGVHQLFDYDNKPEPVNID 540
Db 1814 DDGIMQMKDVKYANGPPDKVTDMDGNVYLQTSYSDLANAKAVGVHQLFDYDNKPEPVNID 1873
Qy 541 PKGNTSIEVADGKSVFVNIDKKNNGPDGEIQOHIIYNGKXYTSFPNDIKQIIDKTLNLIK 600
Db 1874 PKGNTSIEVADGKSVFVNIDKKNNGPDGEIQOHIIYNGKXYTSFPNDIKQIIDKTLNLIK 1933
Qy 601 IYVKDPAKNTTVEKFLNKDQTSYSELKPRVTVTIQNGKEMSTTVSEEDFLPYVKG 660
Db 1934 IYVKDPAKNTTVEKFLNKDQTSYSELKPRVTVTIQNGKEMSTTVSEEDFLPYVKG 1993
Qy 661 LEKGYQPDGMBISGFGKDGAGVYINLSKDTPIKPVFKIIEKKKEENKPTFVSKKKDN 720
Db 1994 LEKGYQPDGMBISGFGKDGAGVYINLSKDTPIKPVFKIIEKKKEENKPTFVSKKKDN 2053
Qy 721 PÖVNHSQLNESHKEDLÖREBHSÖKS DSTYDVTATVLDKNNISSKSTNNPNK 773
Db 2054 PÖVNHSQLNESHKEDLÖREBHSÖKS DSTYDVTATVLDKNNISSKSTNNPNK 2106

RESULT 2

A97942
metalloproteinase (BC 3.4.21.-) A [Imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: A97942
R:Hoekstra, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; Dehoff, B.S.; Ee, R.; LeBlanc, D.C.; Lee, L.N.; Lefkowitz, E.O.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.B.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Balazs, R.H.; Jaskunas, S.R.;
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; M0ID:21429245; PMID:11544234
A:Accession: A97942
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2144 <KUR>
A:Cross-references: UNIPROT:Q8DQF7; GB:AE007317; P1DN:AAK9365.1; P1D:G15458138; GSPDB:C
C:Genetics:
A:Gene: pta
C:Keywords: hydrolase; serine proteinase

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
Best Local Similarity 92.4%; Pred. No. 6,4e-151;
Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;
Qy 1 KLGELASBPKNLGNKESGLKOTTVGEHHQENBSIKSSPTIDNISTIDFENK 60
Db 1339 KLGELASBPKNLGNKESGLKOTTVGEHHQENBSIKSSPTIDNISTIDFENK 1397
Qy 61 DLKGLIKKKPRBYDPTSETGKMEERYDYKDDKGNIIAYDGGTDLLEYETKLDIKSKI 120
Db 1398 DLKGLIKKKPRBYDPTSETGKMEERYDYKDDKGNIIAYDGGTDLLEYETKLDIKSKI 1457
Qy 121 YGVLSPSKDGHEILIGKISVNSKALVYVYGNNTYSIEIKATKYDFHSKTTFPDLVANI 180
Db 1458 YGVLSPSKDGHEILIGKISVNSKALVYVYGNNTYSIEIKATKYDFHSKTTFPDLVANI 1517
Qy 181 IYDGLAPADMDLFPVNDQKAEIKIEMPEKIKETKSEVPYSSYGVNIEIGEGDLSKN 240
Db 1518 IYDGLAPADMDLFPVNDQKAEIKIEMPEKIKETKSEVPYSSYGVNIEIGEGDLSKN 1577
Qy 241 KPDNLTAKMESGKISYDSERKQVYLLKNNIILKSGYALKVTTYNPKGTDMLKGNVYSKEDI 300
Db 1578 KPDNLTAKMESGKISYDSERKQVYLLKNNIILKSGYALKVTTYNPKGTDMLKGNVYSKEDI 1637

Qy 301 AKIOKANPNLRALSETTTIYADSRNVEDGRSTOSGLMSALDGFNIIIRYQVTFPGKNDKGEA 360
Db 1638 AKIOKANPNLRALSETTTIYADSRNVEDGRSTOSGLMSALDGFNIIIRYQVTFPGKNDKGEA 1697
Qy 361 IDKQGLVTDSSSLVLFQKDKKEYTGDBKFNVAIKEDGSMLEPDTKPVNLSMDKXFNFP 420
Db 1698 IDKQGLVTDSSSLVLFQKDKKEYTGDBKFNVAIKEDGSMLEPDTKPVNLSMDKXFNFP 1757
Qy 421 SSKNKTYYNPEFYLGKISDKCGFWMELRVNESVVDNVLIIYGDHIDNTRDNRNKLNVK 480
Db 1758 SSKNKTYYNPEFYLGKISDKCGFWMELRVNESVVDNVLIIYGDHIDNTRDNRNKLNVK 1817
Qy 481 DDGIMQMKDVKYANGPPDKVTDMDGNVYLQTSYSDLANAKAVGVHQLFDYDNKPEPVNID 540
Db 1818 DDGIMQMKDVKYANGPPDKVTDMDGNVYLQTSYSDLANAKAVGVHQLFDYDNKPEPVNID 1877
Qy 541 PKGNTSIEVADGKSVFVNIDKKNNGPDGEIQOHIIYNGKXYTSFPNDIKQIIDKTLNLIK 600
Db 1878 PKGNTSIEVADGKSVFVNIDKKNNGPDGEIQOHIIYNGKXYTSFPNDIKQIIDKTLNLIK 1937
Qy 601 IYVKDPAKNTTVEKFLNKDQTSYSELKPRVTVTIQNGKEMSTTVSEEDFLPYVKG 660
Db 1938 IYVKDPAKNTTVEKFLNKDQTSYSELKPRVTVTIQNGKEMSTTVSEEDFLPYVKG 1997
Qy 661 LEKGYQPDGMBISGFGKDGAGVYINLSKDTPIKPVFKIIEKKKEENKPTFVSKKKDN 720
Db 1998 LEKGYQPDGMBISGFGKDGAGVYINLSKDTPIKPVFKIIEKKKEENKPTFVSKKKDN 2057
Qy 721 PÖVNHSQLNESHKEDLÖREBHSÖKS DSTYDVTATVLDKNNISSKSTNNPNK 773
Db 2058 PÖVNHSQLNESHKEDLÖREBHSÖKS DSTYDVTATVLDKNNISSKSTNNPNK 2110

RESULT 3

ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: T28317
R:Alfonso, C.L.; Tulman, E.R.; Lu, Z.; Oms, B.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A>Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; M0ID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFD>
A:Cross-references: UNIPROT:Q9YVT6; EMBL:AF063866; NID:94049647; P1DN:AAK97677.1; P1D:
C:Genetics:
A>Note: MSV156

Query Match 6.3%; Score 253.5; DB 2; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.0015;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;
Qy 7 ESKFKNLGNKESGLKOTTVGEHHQENBSIK--EKSFTID-NTISTIDFENKDLK 63
Db 196 ESKFKNLGNKESGLKOTTVGEHHQENBSIK--EKSFTID-NTISTIDFENKDLK 255
Qy 64 KLTKKKPRBYDPTSETGKME-----EYDYKIDKGNIIAYDGT--DLEYETE 111
Db 256 KLTKKKPRBYDPTSETGKME-----EYDYKIDKGNIIAYDGT--DLEYETE 315
Qy 112 KLDEIKSKIYGVLSPSKDGHEILIGKISVNSK-----NAKYVYGNNTYSIEIKAT 161
Db 316 KLDEIKSKIYGVLSPSKDGHEILIGKISVNSK-----NAKYVYGNNTYSIEIKAT 367
Qy 162 KIDFHSKTTMFPDLYANINDIYDGLAPAGDMLFVK--DNDQKA-EIKIRMEPKIET-- 216
Db 368 KIDFHSKTTMFPDLYANINDIYDGLAPAGDMLFVK--DNDQKA-EIKIRMEPKIET-- 425
Qy 217 ----KSEY---PVVSYGVNIEIGEGDLSKNPNDLTAKMESGKIYDSERKQVYLLKNNI 269
Db 217 ----KSEY---PVVSYGVNIEIGEGDLSKNPNDLTAKMESGKIYDSERKQVYLLKNNI 269

```

Db      426 KIDEQTEYYKINKINQVNDIIEIKNNNLQKLEENKKIDQTEYYKINKINKEY--NDIIE 483
Qy      270 LRKXVALKATTVNPG-----KTMLENGGVSKEDIQKANKPMLBALSETTIVAD 321
Db      484 LKNNNLQKLEENKINNDKLTGLKNDISNTELEFKMLNIDFKDKSRIAKLN--TETBQ 541
Qy      322 SRN--VEDGRSTOSVL-----MSALDGFNIIRYOVFTPKMDKGAIDKQGLVDSRK 373
Db      542 LRKDLLENINKTNELMKLSDMKSLSL-----QLYDSKKVIL--DGIDKIYNSLKKN- 592
Qy      374 LVLFQKDKAYTGDEKFNVEAIKED--GSMLEFDTKCVNLSMDKNYR--PSKSKLY 427
Db      593 ----DKIDEYFSNTEKFDIYVNIENKFIQNLDSIINKIINDQFKKEYINSKIDSKEIS 648
Qy      428 VRNPEFY-----LRGKISDKGFGNMLERVNMSGVANNLYIGDLIHDNTRD 472
Db      649 TMRPDDIFPAKQIASITNNIENISNKTID-----LNEPITINNESSKEL--IDETRK 698
Qy      473 FNILKLVKDGIDIMQMKDYANGPPDKVTTMDGNV--YLQTVSDLMNA-----V 522
Db      699 YKQGFQ--KIKDAMNTEVKSFR--NLLQKIDISIKSNINELTNAYDIINTKANDLDKLAN 756
Qy      523 GVHVFQFLYDNKRP-----EWNIDPKGNTSIEYAD--GKSVPRINIKNNNGPDGEIQOBI 576
Db      757 GSEBKALYNNASDLLDTIYQKNNDEKVKQLEBYLEKKNQSIETINDIVNN----- 805
Qy      577 YINGEYTSFNDIKQIIDKTINIKIVKDPANNTVKEFPIINKDGEVSB--LKPFRVT 633
Db      806 PI--KELIKPNTB--TKSLN--ELTND--DINDKIPLKYKELANKSTNNLKIYK-- 855
Qy      634 VTIONGKMSSTIVSEEDFPIIPVYKGELEKGYQPDGWEISGPEGKQDAGVYINLSKTP- 692
Db      856 NEINQVNEKLSIVIENTQFINSPLSIEFNQG--SITSH-----INFLANTLA 900
Qy      693 -IKDPFKKIEKKEEENKPTDVSKKKQNPQVNSQLSNESHKEDLQREHSQKDSRTKD 751
Db      901 GINDVNLKNTLKNIMADVTRRGDTNIR-----DEIKNQISSENIKQKENEKE--KD 950
Qy      752 VTATV-----LDKNNISSKSTNN 770
Db      951 LKGLISFNDKLNKNISAGYTEYN 974

RESULT 4
T18444
hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18444
R:Lawson, D.; Bowman, S.; Barrett, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18444
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1650 <LAW>
A:Cross-references: UNIPROT:O77328; EMBL:Z98547; NID:e1325376; PID:e1427940; PIDN:CA8111
C:Genetics:
A:Map position: 3
A:Insertions: 1597/3; 1625/3
A>Note: C0385c

Query Match      6.2%; Score 250; DB 2; Length 1650;
Best Local Similarity 20.8%; Pred. No. 0.004;
Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;

```

```

Qy      130 GHPEILIGIS-----NVSNAKVYGNNTKSI--EIKATKDFHSTMTFEDLYANNIDIV 182
Db      502 DEMELYNQHTDPAFNIENLANKIYF--DIYBGDPKPKKKKKLDHLYTQKKRANNINDL 560
Qy      183 DGLAPAGMRLFVKQNDQKAEIKIRMPBKIK-----ETKSEYPYVSSYGVNIELEG-- 235
Db      561 K-----DHHLNDKETREKQNEIIEEEKNNKIBIEEEKNNKIBIEEEKNNKIBIEEEK 614
Qy      236 ---DLSKRPDNLTNESGKIYSOSEKQYLLKNII-----LR 271
Db      615 KKIIEEEKKKKI--EKEEKKKKIDDEKKNYANDKISHIDNVNCKIIDLALHIEEEK 673
Qy      272 KGYALVTTYPGKTD-----MLBNGVYSKEDIK-----IQKANPML-- 310
Db      674 KTGHEINILYKEIKNEIYQKMLNDENSIMLEHEKKNNTQVANNNLCDTKMLQKKEKILN 733
Qy      311 RALSETTIYADSRVNDGRSTQSVLMSALDGFNIIRYOVFTPKMDKGAIDKQGLVTD 370
Db      734 NDKKKTFLSKSKNI-----TSNVLSKIPG-----TLSTCKLNAATYKTIKCD--VTD 779
Qy      371 SSKVLPFG-----KDDKE--YTGEDKE--NYEA-----IKEDGSMLEFIDTK 407
Db      780 NEKKKYVDHDKKNIKKKKEPINITKGRANAYEIGSBVCNKNVYKQDNNKQVENEK 839
Qy      408 P-----VNLMDKNYF-----NPSKSKNIYYRNPFF 433
Db      840 QGDDNNMVAENKQGENNVIVYNIERSSSPFTFHRKNTSTSTDTCAKNNQKPYHL 899
Qy      434 YLRGKISDKGFMNLEL--VNESVDNY-----LTYGLAHDN----- 469
Db      900 Y--SNQNEKDKKSIPLKNINENIKKNYDKERISTLEKKVFVKQKNVITNNDEKRTSK 958
Qy      470 -TRDENI-----KLN----- 478
Db      959 INDDPNTIYDQKGLKNNPPDLANKIKNETKLEKDSHNSKIQNNLKKKKTNPPTNKK 1018
Qy      479 -VKQGDIMQMKDYKANGPPDKVTTMDGNVYL-----QTK 513
Db      1019 GISSTSISSSTKQNKDGGILEKKNLNLKTYCIRKNNVDSIKLNDKADLYKDKKTS 1078
Qy      514 YSDLNAKAVGVHVFQFLYDNVPRVNIIDPKGNTSIEYADGKSVPR--INDKNNNGPFG 569
Db      1079 FNDIRAKGNNFK-----KRDV--PRKNMMDVTNKKR--VFNPVTLANNYRNN-- 1124
Qy      570 EIOEHIYINGEYTSFNDIKQIIDKTINIKIVKDPANNTVKEFPIINKDGEVSELKP 629
Db      1125 -----YIRSNK--NNVK-----NGMNV-----GIKKIYVLKE--KQKSLHP 1156
Qy      630 HRVYVTIONGKMSSTIVSEEDFPIIPVYKGELEKGYQPDGWEISGPEGKQDAGVYINLSK 689
Db      1157 EGVEADKGLANSVNDKYLEKQGFDDIINEEMEK-----YKNNK--MKYKI--K 1201
Qy      690 DFTIKPVREKIEKKEEENKPTDVSKKKQNPQVNSQLSNESHKEDLQREH--HSQKS 746
Db      1202 SNSIPIPIIKIIRKSNNDNN--DNIRKNDNINSNNINSNDKCLFLSKEDRVHLKKN 1258
Qy      747 DSTKQVATVLDKNNISSKSTNNPNK 773
Db      1259 NIIVNNTMMFRKQSNSCDNTTSLKXK 1285

RESULT 5
T18427
hypothetical protein C035c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrett, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBJ

```

A:Molecule type: DNA

A:Residues: 1-3724 <LAW>

A:Cross-references: UNIPROT:O77320; EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CA3111

C:Genetics:

A:Introns: 307/1; 1545/2

A:Note: C0335C

Query Match 5.8%; Score 235.5; DB 2; Length 3724;
 Best Local Similarity 18.7%; Pred. No. 0.041;
 Matches 192; Conservative 150; Mismatches 312; Indels 371; Gaps 45;

```

QY 5 IAESEFKNLGNGKES--LKK-----DTGVHHHNGNEBSIK-----EKSFT 46
DB 249 IPKQIFENILANKQNDVLRNIIIMLDVNDIMPLHBPDSQNSLNGKGLTGNKKKEKM 308
QY 47 I-----DRNISTRDENDKLKKRREVDPTSEKREEDYKDDG 95
DB 309 IPKGTGYODKEKESLITINQNDKXKCK--KKQSELDSDNISNNLTLSKRYTYT--- 363
QY 96 NIIAYDDGTDLEETEKLDEIKSKIYGVLSPSKDGHEILGKISNVSNAKYVYGNYS 155
DB 364 -----CGMDKETKEDEBQNR-----RNTVVASICVAVETKEKQKKTNNKK 407
QY 156 IEIKATKQDFHSKTYTFDLVYANINDIVGLAPAGDMR--LQYKNDQKKAETIRMP-EK 212
DB 408 ---SVQKODLYEENALD---NL-IIIDGINFDVYTKCKIIDDNNNNENIDNNIYENNK 460
QY 213 IKETSEEPYVSSYGVNY-----ELGSDLSKRPDNTTKMESG 252
DB 461 LKQDQS--VDFSSBKNMLIGVNEGEFEFENIEKLEQERKQKKNQK-----T 512
QY 253 IYSDSEKQOYLLKDNIILEKGYALKVTYTPGKTMLSGNGVSKEDIKIOKAN-PNLR 311
DB 513 IYNNNEBQTDLDNRNI-----NKIESINNNDN---NNNINKEKPKITHTHILNK 562
QY 312 ALSETTYADSRNVEDGSTQSVLMSALDGFNIIRYQVTPYRNDKGAIDQGNLVDS 371
DB 563 SISKIIRKSPSRDK-----IKLYTNNKGBSTFELKLELITNN 604
QY 372 SKLVYFGDDKQYTGDEKFNVAIKEDGMLFDTRPVNLSMDKQYFNSKSN----- 424
DB 605 KVNVT--BEDIIGSNEDRYIHLVX-----NLKEDANEYNNDKKNKNNKTK 650
QY 425 ---KIYVRN-----PEFYLRKISDKGFMNLEVN---ESVVDNYLYIGDLHIDTSD 472
DB 651 ILKSNVYLENERLTJELKLRG---NNIFKQKRNKSGVYIINEIQNEENKINDIOD 707
QY 473 FNIR-----LNVFD---GDIMWGMCDYKANGPDPKVTMD----- 505
DB 708 GNISKQKLIQSSSRNTDPTFNIDISLNDLEKRRKKSQHPIDMLVKADKKEISENIK 767
QY 506 -----GNVYLGQ-----GYSDLAKAKAVGVH----- 526
DB 768 ICDNNNNIYBESINNIIYBESINNIIYBESINNIIYBESINNIIYBESINNIIYB 827
QY 527 -----QFLVNYKPEVNID-----PKGNTSIE-----Y 549
DB 828 NNNNIYBGINIKI CODNILENNKIKTNDIVQVEENNSIEKQBLMSLANKOINNYNMF 887
QY 550 ADGKSVVRN-----INDKRNNGFDGRIQ----- 573
DB 888 KENVDIPIFKIKRBSLAKIDKNIKONNNDDEYIMNPFENDEPIINHKEIITNKEIDPLEI 947
QY 574 -----OHYINGKETS---FNDIKQIIDLTKINIKIVQFAIRTYVKEFILNDQTS 623
DB 948 NTQNEFIENLDIKKKGTINDHPTNDADKPYE-NA-KILANKMGKKQKQKFPKTBETRS 1005
QY 624 VSEIK-----PRVYVTVTQNGKMSSTIYSE 649
DB 1006 LQSHIKIKYKNGKEGKDKNNEBKNIKLYDENQVSVLSYDKIKEDIODIHSIQNIDC 1065
QY 650 BDFILPVYKGBLEKGYQFDGWEISGFEGKQKAGVYINISKOFFIPVFKIKKEKEENK 709

```

```

DB 1066 NN-IBQINENSKGVRISGTDN---ENKND-----MENKNDMEKK 1102
QY 710 PTFDYSKKQNDPOVNH-SQLESREKEDIQREHESQKSDSTQVATVLDKNNISKSTT 768
DB 1103 N--DMEKKNIDIEKKDMEKKNDMEKKNDMEKKNDMEKKNDME-----NENMENKSDI 1153
QY 769 NNPNK 773
DB 1154 ENENK 1158

```

RESULT 6

T28677 rhoptry protein - Plasmodium yoelii

C:Species: Plasmodium yoelii

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T28677; C45521

R:Keen, J.; Simha, K.; Brown, K.; Holder, A.

Mol. Biochem. Parasitol. 65, 171-177, 1994

A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.

A:Reference number: Z20508; MUID:95021522; PMID:7935623

A:Accession: T28677

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2269 <KE2>

A:Cross-references: UNIPROT:O26223; EMBL:L27838; NID:9457145; PID:9457146; PIDN:AAA2130

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990

A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple

A:Reference number: A45521; MUID:91101660; PMID:2270106

A:Accession: C45521

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 2131-2269 <KE2>

A:Cross-references: GB:M34283

Query Match 5.8%; Score 231.5; DB 2; Length 2269;
 Best Local Similarity 19.7%; Pred. No. 0.032;
 Matches 181; Conservative 147; Mismatches 330; Indels 259; Gaps 37;

```

QY 37 ESIKESSTIDRINSTRDFENK-DLKLTKKREVDPTSEKQKMEY----- 87
DB 109 EQLKLELNKKIIDIILAKI EYVNTVELEKKEIKNNAYIDELANQSPYKTYGIEKNTIY 168
QY 88 -----DYKDDKGNIIAYDDGTDLEETEKLDEIKSKIYGVLSPSKQGF 132
DB 169 NTIKSYFDQIYRGDIDPTNELSIVKEDPIDIEDKT-KLENLSKIDNVDKIQMEI 227
QY 133 E-ILKISVSKNAFVYGNNTYSIEIKATKYDFHSKTYTFDLVYANINDIVGLAPAGDM 191
DB 228 ETVKSHLNNIETNNKL-----PWTILEIKYIYDEISK---ELANKLEDFKVK---EKEL 276
QY 192 RLFPVQNDQK---AEIKIRPE-----KIKETSEEPYVSSGVNIIHGE 234
DB 277 SNKISDYDKRQQLSEYKSKMLEIRHNYSQTNVDNTKEBEKKQYDKNSEHMTTIPNNE 336
QY 235 GDLSEK-----NKEDNITKMS-----GKIYSD---SEKQYTL-LKDN---ILKGYAL 276
DB 337 DEISKIIEVTKMOBILSKWATYIDPNKIKYETVNSHESQTBELTDKIKAVSDKELK 396
QY 332 SNVVDGSGTQSVLMSALDGFNIIRYQVPT-----FKNN-----DKGEAL 361
DB 457 NQNIKTVKETSIDISYIERPE---QITLGKQTLLENKFTFESLNHNEANNELIKYFS 512
QY 362 DKQGNLVYDSSGLVYFGDDKQYTGDEKFNVAIKEDGMLFDTRPVNLSMDKQYFNS 421
DB 513 DLKANVLGINENMLNQGTEKTEKTFND-----IKKNHINIBESIKIEIKIHASIVNIS 566
QY 422 KSNKIYVRNPEFYLRKISDKGFMNLEVNESVVDNYLYIGDLHIDTSDRDNITLANK- 480

```



```

Db      567 EETEREIGINIESLATTYFER-----VKENVTNLIKIKELKHYPFSDGKGNITXY 618
Qy      481 -----DGDMDGMK--DYKANGPDP-----VYTMDDGNV-----YLQYSDLNAK 520
Db      619 TDKIKCINDDITMAVSQOQIDOHINGLDDIOKSESYSVEKGEQINLEKYSNTEISNDVYE 678
Qy      521 AVGVHYOFLYDNVKEVNI-----DPRKNTSIEYADKSVVFNN-----560
Db      679 GIKKKQOQIIVTKIDKKNIYEBINKLSIKIEKONTSLKVKQDINSYGCNGLNPLFE 738
Qy      561 --DYKANGPDPGEIQOHYYIN-----GKETYSPNDIKQIIDXTLN-IKI-----601
Db      739 QIDEREKKKAEKNTIKSMKAVYIDLDNKKKKSQEIETEMDIDKMDINKEMELKISHDDKKC 798
Qy      602 -----VYKDPARTTYVETFLNMDTGEVSLKPKRPVYVTLQNGKB 641
Db      799 HDKRNKHEKNTISDIYDKSSKTIQDPSRESIDND-INKKLQKNVSSQNHNSDIN-----851
Qy      642 MSTIVSEEDPILPVYKGELEKGYQPDGMBISGFEKGDAGVYINLSKDTPIKVPFKLIE 701
Db      852 -----QCLMEVANIYNI LKLNKIKKIIDVVK 877
Qy      702 EKKEENKPTDVSQKQDNPOVNSQLAKSHKEDLQREHSQKSDST---KDV-----T 753
Db      878 EYTSIEIK-----NKKINDELNNSEKVIKKIIEGLSLKCRKSNSTLDDKDIDECIKN 932
Qy      754 ATVLDKNNISKSTNN 770
Db      933 INVLAKKILNEETNITN 949

```

RESULT 7

228676
rhoptry protein - Plasmodium yoelii (fragment)

C/Species: Plasmodium yoelii
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T28676, A45521
R/Sinha, K.A.; Keen, J.K.; Ogum, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A/Title: Comparison of two members of a multigene family coding for high-molecular mass
A/Reference number: 220507; MUID:97077455; PMID:8920022
A/Accession: T28676
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2401 <SIN>
A/Cross-references: UNIPROT:Q26216; EMBL:P36927; NID:g1041784; PID:g1041785; PIDN:AA8412
R/Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A/Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A/Reference number: A45521; MUID:91101660; PMID:2270106
A/Accession: A45521
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 2260-2401 <KE>
A/Cross-references: GB:M34281

Query Match 5.7%; Score 230; DB 2; Length 2401;
Best local similarity 21.0%; Pred. No. 0.04;
Matches 187; Conservative 138; Mismatches 292; Indels 274; Gaps 46;
Qy 2 LGIABSKPKNLGNGKESLKKDVTGVA-HHHQENESIKKSSFTIDRNISTIRDFANK 60
Db 589 IDKIYTKFBNILTDKTELETFTGLSLNNHNSNN---KELLYFYDLKAKLGRKKN- 644
Qy 61 DLKGLIKKKPREVDFTSETGKMEYDYKYDDKNILTA-----YDGTGLEVET 110
Db 645 ---MLTKQFNKEK-----KAVEDIKKKQVNDINKTYSNIEITTYTSITNNBTENBI 693
Qy 111 EK-----LDEIKSKTYGVLSPSKDGHPHLLKISINVSNNAKVYVGN 151
Db 694 GKSLBLNTYVLEKVKYANVTMLNBIKELT-----KQYDQDPGK-----EKNTKYPPDN 742

```

Qy      152 NYKS-IEIKATKYDHFSTMTFEDLYAN---INDI---VDGLAFAGMRELKYQNDQKKA 203
Db      743 KIKNDIDTLNQGIDKSIETLT--EIKKSENHIDEIKQIDIKKKKPPNKTMF--NBDPKEI 799
Qy      204 EIKI-RMEKIKETSEKPYVSYGNVLELGGD---LSKNDPLATMES-GKIY-----254
Db      800 EKKIENIYEKIDKKNKIYKIDKLANESIKIENDTSLKELKNNINLSYKSLGNLPLQOI 859
Qy      255 -SDSEKQOYLAK-----DNI-----ILKGYA 275
Db      860 DEKKKKAHTIKAMAWAYIDLDNIIKKSQOIEKENINMDIKMDIHKKMKALNISHDDVK 919
Qy      276 LKVTY--YNGKTDMLKNGVYSKEDIKIQKAPNLALSETTYIYADSRNVEDGKSTOS 333
Db      920 IYHTTSKHHBEKISDIRKNSLKTIQDFSESYINDIKKELEKNVLESQNNNTDINQ---975
Qy      334 VLMSLDC-FNIRQVTFPKNDGELADDOGANVYTSSKLYLFKGDQKXYTGR-----387
Db      976 -YLSKIENTYNI-----KLNKIKKIIDK-----VKETTDIEKKN 1010
Qy      388 DEFNVE-----AIREDSMLPIDTKPVNLSDKXNYFNPSSKNKIYVRNPEFYLRGK 438
Db      1011 KKNLNLSENSEKIIYQLAKENSLKECQSK-IKSTIDNIVYSECIGN---ITLKYTYI---1063
Qy      439 ISDKGFNWEKLVNESVDNY-LIYGLHIDNTDIPNIKLVYQGDINDMGKQDYKANGP 497
Db      1064 VIEKKNINITYFNAAEYQNVNSLNFNNIEMADTKSQYI-LNIX-----KNNG- 1109
Qy      498 PDKVTMDGNYVLYQYSGDLNAKNGVAYQFLYDNVKEPVNIDPRKNTSIEYADKSVYF 557
Db      1110 -----TNNTDINIKELKHK-----KSNVYDENGKRTQELKKNKEL-- 1147
Qy      558 NINDKRNNGPDEIQEHOIYNGKEYTSFNDIKQIIDTFLNI-KIYVDF--ARYTYVKS 614
Db      1148 -----FEKYGEVYVLLN-KYAV--ELKNKFDKTNVYSEQIIEIKDPAHPTFSQ 1195
Qy      615 FLNMDTGEVSLKXHYVTVTLQNGKESSTIVSEEDTILPYKGELEKGYQPDGMEISG 674
Db      1196 --ADSEKKMEIKNEQRIEDEVAKNNKSN-----KALD-----1229
Qy      675 FEKKKADAGVYINLSKDTPIKVPFKLIE---KKEENKPTDVSQKQDNPOV-----723
Db      1230 -----IQLSVEEP-KIKFLKIKDLRTKSDCLKKETDIETKISNLSIDTQETKLI 1278
Qy      724 -NHSQLANSHKREDLQREHSQKSDSTQVATVLDKNNISKSTNNPNK 773
Db      1279 ENKNILNLTLEKLSLKNQKKNIDQKK-----LDEVNSKIKNIESNVNQ 1324

```

RESULT 8

TI8440
hypothetical protein C0425W - malaria parasite (Plasmodium falciparum)

C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: TI8440
R/Lawson, D.; Bowman, S.; Barrett, B.
Submitted to the EMBL Data Library, August 1997
A/Reference number: Z18935
A/Accession: TI8440
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-4550 <LAW>
A/Cross-references: UNIPROT:O77336; EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB
A/Genetics:
A/Map position: 3
A/Note: C0425W

Query Match 5.7%; Score 228; DB 2; Length 4550;
Best local similarity 21.6%; Pred. No. 0.11;
Matches 193; Conservative 119; Mismatches 347; Indels 236; Gaps 43;
Qy 5 IABSKPKNLGNGKE---GLKKDDTGV-----HHHQENESIKKSSFTIDRNISTIRDF 56

```

Db      746 IESRSISGNNHEIYGASIKCAQNGVRNKLQGNHNDIKDKDNKKOG---DKKDNKCD 802
Qy      57 FENKQIKKLIKKEFREVDPTSETGKRMREYDYTKDDKNIAYDDGTDLEIF----- 109
Db      803 GQKQDDKKGDDKRYDDGDD-----KYDDDDKRYDDDDDI--YDDDDDFNFHDSSTSK 853
Qy      110 -----TEKLDEIK-SKIYGLSPSKDGHFELIGLIS-NVSNNAVYVYG 150
Db      854 RLSTNDLSLKQDQKGSILKQEDTLKCSKNTNVOEKESRMQCMANTNCCVARE-KNYIF 912
Qy      151 NNYKSIEIKATKYDFHAKTMTPTDLYANINDIYDGLAFAGD-----MLTF-----VKDN 198
Db      913 NDYMHINNTLSDHEMNINMKSPGVNNTLFDVDVHFHKGDFIMNLITLKIKEKIQVLYQ 972
Qy      199 DQKAEIKIRMEBEIKETSEYVYSSYGNV-----ILAGSDLS-----KN 240
Db      973 NKQVCVCFCKTGTGK-TYCMPEPSCSTYTHIYCYNRYKMOHVAKKKNDIOLCASQKSRK 1031
Qy      241 KPDNLTYKNE-SGKIYSDSEKQOYLKQNIILKGYALKVTTYNPQKTDMLBGNVYSKE 298
Db      1032 QRAASVYKRGAGSKQDQSRSLDRKQDQVTRRSNKTESANDSDGKN--NGDDKKKGD 1089
Qy      299 DIAKI QKANPRLALSETTIYADSRNVEDGRSTQSYLMSALDGFNIIRYQVPTFKMNDK 358
Db      1090 NIGKNNNNNNNNDDSDNNNGDDSNNGDDSNNG--DGNNTINGDNNNINGDGN 1143
Qy      359 EALDDQGNLYT-----DSKLYLFRKQDEYTGSEKFN 392
Db      1144 NNINEDGNNNDCLBEKQDQNOKRRKRRKRRNNNDPTSLVLANKDSROKKRKYG- 1202
Qy      393 EALIKEDGSLFIDTYKPNLSMDKNYFNPESKSNKIYVRNPEFYLRGCI-SDKGFNMELRY 451
Db      1203 ---KDH-----DQKDSPSKINNRSKN-----KVKQDKVGN-----I 1234
Qy      452 NESVDNTIYIGDLHIDTRDPTNLIKLVKQD---IMQMKDYTFANGPDKVTDMDGN 507
Db      1235 NDDKINN-----NTNDNNKNNNI--NGDNKNNNINGDNNKNNING--DKI---NNN 1278
Qy      508 VYLGQGYDLNAKAVGVH-QFLYNVKEVNIIDPKGNTSIEVADKSVVFNINDK----- 562
Db      1279 INGDKINNNTINDCKISKNYHSHNYDN--EITHEQKNNFN-KKNTKRYKRSINDKEND 1335
Qy      563 ---RNGFPGS-----IOEOHIYINKEYS-----FNDIKQIIDLTKIIVKQDPA 607
Db      1336 ILNNKNDPTDKKLNKYNFLMEYQKIISSDKITSISNNKDIKNID-----IKDK 1387
Qy      608 RNTTYKKEPILANKQGEVSELPKPHRYVTIYQNGKMSSTIVSEBDPIIPYKGLBEGYQF 667
Db      1388 DIKNIKDI---KDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDI 1424
Qy      668 DQWEISGFEKGDAGV-----INLSKDT--FLKPVFKIEE--KKEENKPTFDVSKKK 718
Db      1425 ---IKDIKDVNDVSSVHLNKCYNKSDTQDPCPSDDKNIINNIKKKKNINIKTDQVY--- 1477
Qy      719 DNPQVNSQLMESHKEDLQREBSQKSD-STQDVATATVLDKNNTSSKSTTNPN 772
Db      1478 -NPLVDSSSSSTDCNRYKKEKAVYKIGMINKIKLHNNNNNNNNNNNNNNNNNNNN 1531

```

RESULT 9

T18429
 hypotheical protein C0345w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T18429
 R:Lawson, D.; Bowman, S.; Bartrell, B.
 Submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18429
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1711 <LAW>
 A:Cross-references: UNIPROT:O77322; EMBL:Z89547; NID:e1325376; PID:e1325381; PIDN:CAB11J

```

C:Genetics:
A:Note: PFC0345w
Query Match      5.5%; Score 221.5; DB 2; Length 1711;
Best Local Similarity 19.7%; Pred. No. 0.059;
Matches 160; Conservative 134; Mismatches 274; Indels 243; Gaps 37;

```

```

Qy      29 EHHQENESIEKES-----FTIDRNIISTIDFENKDLK-----KLKKGREVD 74
Db      984 DNNNDNNNNNNYDSSSSNNHNYIILTNDRKLN-MDNFINNNLEINNNSQNKVEIKNEYIN 1042
Qy      75 DF-TSETGKRMREYDYKQDQNIILAYDDGTDLETEKDLDEIKETIYGLVSPSKDGHFE 133
Db      1043 NVKLTITS-----NFGSNNTNSKDE-----NHISRSKCKEDTILNLSKRSYSYE 1086
Qy      134 ILGKI--SNVSKQAKVYYGANNYSIEIKATKYDFHAKTMTPTDLYANINDIYDGLAFAGDM 191
Db      1087 YNNKILQSTNSKSLANGAYENNLPSGKKKKK-----GVLYKDI-BHINDIQD--KYPEDL 1138
Qy      192 RL-----FYKDDQKAEIKIRMEBEIKETSEYVYSS--YGNVIELGSDLSKRP 243
Db      1139 NINCVRKYVTENEKRL-----LPLEL-----EYNLVSSDEKFG-----LAKI KND 1179
Qy      244 NLTWBSGKIYSDSEKQOYLKQNIILKGYALKVTTYNPQKTDMLBGNVYSKEDIK 303
Db      1180 N-----NITYKQHYHNLVYDN--QKGHILEPTNK-----VSI 1212
Qy      304 QKANPRLALSETTIYADSRNVEDGRSTQSYLMSALDGFNIIRYQVPTFKMNDK--GEAI 361
Db      1213 QR-NNNINSVIRKTHYEVKKNKDKRN-----YDNFTCDKKKKIYVNI 1255
Qy      362 DQDGNLYTSSKLYVLRKQDEYTG-----BDKFNVAIKEDGSLFIDTYKPNVL-- 411
Db      1256 NSDKDIYHNN---IITTKKEKGIQNIHLNRDKDITNELKLDQVKEFLDFFQSYID 1312
Qy      412 ---SMDKNYFNPESKSNKIYVRNPEFYLRGKISDGGFNMELRVNESVVDNYLYIGDLHIDN 469
Db      1313 CNKCKENILMNTNKK-----EDHQIIDVAK-IFN---ETMTIMDNKKIYDDGNVHE 1362
Qy      470 TRDPTNLIKLVKQDITMDQMKQDYKANGPDKTDMD--GNVYLGQGYDLNAKAVGVH 527
Db      1363 KCTHNDVITHHMDILSTSIKNNBETFLIDTYQOKRIRIDYVNR----- 1407
Qy      528 FLYDNKPEVNIIDPKGNTSIEVADKSVVFNINDKNNG-----PDGEIOEHIYINKE 583
Db      1408 ---INLQEDDDDDNNNN-----NNNNNNNNKILIFETTKDQDMLAHKNNNL 1452
Qy      584 TSFNDIKQIIDLTKIIVKQDFARNTYKKEFLNKGQGEVSELPKPHRYVTIYQNGKMS 643
Db      1453 EGTBEESDPIEKKNYKIKKQKESYHKIDESILSNKNNKVSIL-----LINNKDSS 1505
Qy      644 STTVSEEDPIIPYKGLBEGYQFQDQWEISGFEKGDAGVYINLSQDTIKVPFKIEEK 703
Db      1506 SV-----DNNKNNNNKNNNNNN-----NNNN 1527
Qy      704 KEENKPTFDVSKKKQNPQVNSQLMESHKED-----LQREBSQKSDTQDVATATV 756
Db      1528 NKKNNKNNNNDSFSDNNLINDNNNNNNNNDSFSDNNLINDNNNNNNNNNNNNNNNNNN 1587

```

RESULT 10

A24594
 probable major surface antigen (83k, 19k, 42k) precursor - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jul-2004
 C:Accession: A24594
 R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, Nature 317, 270-273, 1985
 A:Title: Primary structure of the precursor to the three major surface antigens of Plas

A:Reference number: A24594; MUID:86014355; PMID:2995820

A:Accession: A24594

A:Molecule type: DNA

A:Residues: 1-1640 <HOL>

A:Cross-references: UNIPROT:P04933

C:Superfamily: major merozoite surface antigen

C:Keywords: surface antigen

Query Match 5.4%; Score 219; DB 2; Length 1640;

Best Local Similarity 18.6%; Pred. No. 0.071; Mismatches 340; Indels 228; Gaps 39;

Matches 167; Conservative 164;

```

23 KDTTG-VEHHHGENESIK-----EKSFTIDRNISTIDPENDKLKKKKKFEVVD 75
232 KDNVGMEDYIKNNKKTENINELIESKTTIDKNNAKKEE-----KKKLYQAY 283
76 FTSETGRMEEDYDYDKNIIAYDGTDLAYETEKLEDEIKSIKYGLVSPKDGFEIL 135
284 DLSYNNKQLEB-----AHNLSV-----LE---KRIDTLK-----KRNKKEIL 319
136 GKSIVSKAKAVYVGNVNNKSIKIKATKDYH-----SKMTPTDLYANINDIVGLARA 188
320 DKINIKKPPANSNGTPTLIDKNNKKEHEKEIKELAKTIKFNIDSLPTDPL----- 373
189 GDMELFVNDNDQKAKIKRMPEKIKETSEYP-----YVSYGVN-----IELGE-GDLSEK 239
374 -ELEYLELREKN-KNIDISAKVETKSTENEPNGVTYPLSLNDINNALNELNSFGDLI- 430
240 NKPDNLTYMESGKIYSDSEKQYL--LKNIDIL-----RKGYALKVTTYNPGKTDYL 289
431 -NPFDTYKPSKNITLYDNEKKYINIEIKIKIEKKKESDKSYEDRSKSLN---DIT 485
290 EG-----NGVYSKE-----DLAKIQQA-----NPNLPALETTIYADSRVNEGDSSTOSV 334
486 KEYEKLENEIYDSKFNANNIDLTNEFKMGKRYSYKVEKLTHTNTPTAFSYENSKNLEKLT 545
335 LMSALDGP---NI-----RYQVPTFKNDKGBALDKDGNLVTDSKVLPLFGKDKEXT 385
546 ALKTMEDYSLNIVVEKEKLYKNLISKLENEIETLVE--NIKDBEQLP-----EKKIT 598
386 GEDKFNVAIKEDGMLPIDTYKPNVLSMDKATPNPSKSKIKIYVRPEFLRGKISDGKGF 445
599 KDNKPPDKELFVSDIVKVOQVLL--MNKIDELKKTQLILKQVE--LKNHIVPNVSY 653
446 NMELRVNBSVDNLYLYGDLHIDNTDFENIKL-----NVK----- 480
654 KQENKQEB---PYLLIVLKKGEIDKAKVMPKVESLNEBKKNIKTQGGSDNSEPSTGEI 709
481 -----DGDIMDGKMDYK-----ANGFPDXY 501
710 TGAATTKGQAGSALBEGDSVQAQAEQQAQPPVPVPAKAAQVPTPPAPVNNKTEV 769
502 TDMGQNVLYQGYSLDNL-----KAVGHYQPLVNDVAKPEVNIIDKGNISIRYADGKSV 556
770 SKLD---VLEKLYQPLNTSYICHKYLIVSHSTNNKILIKQYKITYEBSKSSCCPLDIL 826
557 FNIN-----DKRNGFPDQ---EIOEGHIYINGKEYTSFNDIKQIIDDITLNIKIYV 603
827 FNIGNINIVNYSWMPISLNSLSQLEMEIYEEKVNCVLYLKDNDKIKNILEBAKAVSYISV 886
604 KDPANNTVKEPILANKQGEVSELPHRVTVTIQNGKEMSSTIVSEBDFILPVYQGELEK 663
887 KTLSSSSMQPLSLTPQDPEVSAANDTSHSTNNLSLKLFEHILS-----LGNKNG 937
664 GYQPDGMEISGFEK-----KQAGYINLSKOTFKIPVFKIKBEKKEENKPTP--D 713
938 IYQ---ELIGQKSENFYKILKDSQDTFTNBSFTNPFASKADDDINSLANDESKKRLLED 993
714 VSKKDNQVNSQLANESHKEDLQREBSQKSDSTQVATATVLDKNNISSK-STTNP 771
994 INKLKKTQLSLPDLNKKYKLEKRLFDKKKTYGKTKMQIKKLTILKEQLESLNLSLNP 1052

```

RESULT 11

SAZOKI

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #ext_change 09-Jun-2000

C:Accession: A25120

R:MacKay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.

A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium f

A:Reference number: A91030; MUID:86136024; PMID:3004972

A:Accession: A25120

A:Molecule type: DNA

A:Residues: 1-1631 <MAC>

C:Comment: The merozoite stages of different strains have strain-specific surface ant

C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane pr

F:1-19/Domain: signal sequence #status predicted <Sig>

F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>

F:67-84/Region: 3-residue repeats (S-G-T/P)

F:1614-1631/Domain: membrane anchor #status predicted <MBN>

F:97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (asn

Query Match 5.4%; Score 218; DB 1; Length 1631;

Best Local Similarity 18.1%; Pred. No. 0.078; Mismatches 339; Indels 240; Gaps 37;

Matches 164; Conservative 162;

```

23 KDTTG-VEHHHGENESIK-----EKSFTIDRNISTIDPENDKLKKKKKFEVVD 69
223 KDNVGMEDYIKNNKKTENINELIESKTTIDKNNAKKEEKKLYQAYQDLYFNKQ 282
70 FREVDFTSETGRMEEDYDYDKNIIAYDGTDLAYETEKLEDEIKSIKYGLVSPKDG 129
283 LEBANLISVLEKID-----TLKKNENIK----- 307
130 GHEFLIKSIIVSKAKAVYVGNVNNKSIKIKATKDYH-----SKMTPTDLYANINDIV 182
308 ---ELIDKINEIKNPPANSNGTPTLIDKNNKKEHEKEIKELAKTIKFNIDSLPTDPL 364
183 DGLAPAGDRLFVNDNDQKAKIKRMPEKIKETSEYP-----YVSYGVN-----IELGE 234
365 -----ELEYLELREKN-KNIDISAKVETKSTENEPNGVTYPLSLNDINNALNELNS 416
235 -GDLKKNKPDNLTYMESGKIYSDSEKQYL--LKNIDIL-----RKGYALKVTTYNPG 283
417 PGDLI--NPFDTYKPSKNITLYDNEKKYINIEIKIKIEKKKESDKSYEDRSKSLN- 473
284 GKTMLBEG-----NGVYSKE-----DLAKIQQA-----NPNLPALETTIYADSRVNEG 328
474 ---DITKEYEKLENEIYDSKFNANNIDLTNEFKMGKRYSYKVEKLTHTNTPTAFSYENSKN 530
329 RSTQVLSALDGP---NI-----RYQVPTFKNDKGBALDKDGNLVTDSKVLPLFGK 379
531 LEKLTAKTMEDYSLNIVVEKEKLYKNLISKLENEIETLVE--NIKDBEQLP----- 584
380 DDKBYTGEDKFNVAIKEDGMLPIDTYKPNVLSMDKATPNPSKSKIKIYVRPEFLRGK 439
585 -EKKITKQENKQEB---PYLLIVLKKGEIDKAKVMPKVESLNEBKKNIKTQGGSDNSEP 694
440 SPKGFNMEKRVNBSVDNLYLYGDLHIDNTDFENIKL-----NVK----- 480
639 HVPNSYKQENKQEB---PYLLIVLKKGEIDKAKVMPKVESLNEBKKNIKTQGGSDNSEP 694
481 -----DGDIMDGKMDYK-----ANGFPDXY 495
695 STEGEITGAATTKGQAGSALBEGDSVQAQAEQQAQPPVPVPAKAAQVPTPPAPV 754
496 GPPDVTDMGQNVLYQGYSLDNL-----KAVGHYQPLVNDVAKPEVNIIDKGNISIRY 550
755 NKTENVSKLD---VLEKLYEFLNTSYICHKYLIVSHSTNNKILIKQYKITYEBSKSS 811
551 DGKSVVFNIN-----DKRNGFPDQ---EIOEGHIYINGKEYTSFNDIKQIIDDITL 597

```

Db 812 DPLDLLFNIGNNIPVWMSFDSLNNSLSQLFMEIYKEMWVGNLYKLDKNDKDKNLLBEAK 871
 Qy 598 NIKIVVADPARNTTYKBEPLANKOTCEVSELRHRTVTITIQNGKMSSTIVSEEDPILPVY 657
 Db 872 KVTSTSVKLTSSSSMQLSLTPQDKPEVSNANDTSHSTINLNSIKLPEINILS----- 922
 Qy 658 KGBLEKGYOPDMWISGFEKG-----KDAGVYINLSKOTFIKVPFKIEKKKEBENK 709
 Db 923 LGKNKNTYQ-----ELIQKSSSENFYEKLIKDSPTFNBSFTNFVSKADINDSLNDESK 978
 Qy 710 PTF--DVSKKDPQVNSQNLNESHKEDLQREHSQSDSTKDVATVLDKNINSSK-S 766
 Db 979 KLEEDINMLKKTQLSLFQLYNRYKLIKLERLPDKKTVGRYQVQIKQLTLKQLESKLN 1038
 Qy 767 TTNP 771
 Db 1039 SLNMP 1043

RESULT 12

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (sc
 S05603
 N/Alternate names: gp195, surface antigen
 C/Species: Plasmodium falciparum
 C/Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
 C/Accession: S05603; S04850
 R/Miyer, P.J.
 submitted to the EMBL Data Library, April 1989
 A/Accession number: S05603
 A/Accession: S05603
 A/Molecule type: mRNA
 A/Residues: 1-1639 <MTL>
 A/Cross-references: UNIPROT:P04933; EMBL:X15063; NID:g9896; PIDN:CAJ3163.1; PID:g9897
 R/Miyer, P.J.
 Nucleic Acids Res. 17, 5401, 1989
 A/Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plas
 A/Reference number: S04850; MUID:89345116; PMID:2668887
 A/Accession: S04850
 A/Molecule type: mRNA
 A/Residues: 1504-1639 <MTL2>
 A/Cross-references: EMBL:X15063
 C/Superfamily: major merozoite surface antigen
 C/Keywords: glycoprotein; merozoite; surface antigen
 F/1-19/Domain: signal sequence #status predicted <Sig>
 F/20-1639/Product: major merozoite surface antigen #status predicted <Mat>

Query Match 5.4%; Score 218; DB 2; Length 1639;

Best Local Similarity 18.6%; Pred. No. 0.079;

Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;
 Qy 23 KDTTG-VEHHQENESIK-----EKSFTIDRNISTIRDPENKDKLKIKKGFREVD 75
 Db 232 KDNVGRMEDYIKKNGKTIENINELBESKCTIDKRNKKEB-----KKGLYQAY 283
 Qy 76 FTSETGKMEEDYDYDKNIIAYDGDGLFETEKLDIISKYGVLSPSKDGFEIL 135
 Db 284 DLSINRQLES-----AHNLISV-----LR--KRIDTLK-----KRNKIKEL 319
 Qy 136 GKISVSNNAKYVGNNTKSIKIKATKYDF-----SKMTFQDLYANINDIYDGLAFA 188
 Db 320 DKIMIKNPPRANSQNTPTLNDKNNKLEHSEKIKETIAKTIKPNIDSLFTPL----- 373
 Qy 189 GDMRLFVNDQOKAKIIRMEPKIKETSKRYP-----YVSYGAV-----IRLGE-GDLSK 239
 Db 374 -ELEYYLBEKN-KNIDISAKVETKESTBENETPGVTYPLSYNDINNALNELNSFGDIL- 430
 Qy 240 NKPDULTYTGSGKIKYSDEKQYL--LKDNIL-----RKGLAKATYTYNPGTMDL 289
 Db 431 -NPDYTEPSKNITYDNERKFPINBIKIKIKKKKIESDKSYEDRKSLN-----DIT 485
 Qy 290 EG-----NGVYSKE-----DIKIQKA-----NPNILALSETTIYADSRNVEDGSGTOSV 334
 Db 486 KEYEKLALBIYDSKNNNIDLTNFRKMKRYSYVKEKLTTHNTYFASYSNNGNLEKLT 545

Qy 335 LMSALDGF-----NII-----RYVTFPMQNDKGEALIDKQNLWTDSXYLFGKDXEY 385
 Db 546 ALKIMEDSLNRITVVERKEIKTKNLIISKIENIEFTLVE--NKKQBEOQLP-----EKKT 598
 Qy 386 GEDKFNVAIKEDGSMLFIDTFVNLSDMKYFNFSKSNKIVYRNPEFLRGKISDGKGF 445
 Db 599 KQENKPDKILLESVDIVAVQVQKVL-----MNRIDELKTKQLIKKVE--LKNINHYPNISY 653
 Qy 446 NMELVNESVDNNTLIYGDMLIDNTRDNRIK-----NVK----- 480
 Db 654 KOENKQK-----PYULIVLKEKIDKLVMPKVESLJNEBKKNIKTEGSDNSBPSTEGH 709
 Qy 481 -----DDGIMQMGKDYK-----ANGPDKY 501
 Db 710 TGOATTRGQAGSLBEGDSVQAOKQOKOPPPVPPVPAKAQVPTPPPAVNNKTAV 769
 Qy 502 TDMQGNVYLQGYSLDLA-----KAVGVHYOFLYDNVYKPEVNIIDPKGNTSIEYADGKSV 556
 Db 770 SKLD-----YLRKYERLNTSYICHKYLIVSHSTNNEKILKQYKITBEESKLSSCPDLIL 826
 Qy 557 FNIN-----DKRNNGFDG--EIOQHIIYINGKAYTSFNDIKQIIDTKLNTKIYV 603
 Db 827 FNIGNNIPVWMSFDSLNNSLSQLFMEIYKEMWVGNLYKLDKNDKDKNLLBEAKKVSTSV 886
 Qy 604 KDFARNTTYKBEPLANKOTCEVSELRHRTVTITIQNGKMSSTIVSEEDPILPVYKGELEK 663
 Db 887 KTLSSSMQPLSLTPQDKPEVSNANDTSHSTINLSIKLPEINILS-----LQKNKA 937
 Qy 664 GYOPDGMWISGFEKG-----KDAGVYINLSKOTFIKVPFKIEKKKEBENKPTF--D 713
 Db 938 IYQ-----ELIQKSSSENFYEKLIKDSPTFNBSFTNFVSKADINDSLNDESKRKLEED 993
 Qy 714 VSKKDNQVNSQNLNESHKEDLQREHSQSDSTKDVATVLDKNINSSK-STTNP 771
 Db 994 INKLKKTQLSLFQDLYNRYKLIKLERLPDKKTVGRYQVQIKQLTLKQLESKLNINMP 1052

RESULT 13

lipoprotein (imported) - Mycoplasma pulmonis (strain UAB CT1P)
 C/Species: Mycoplasma pulmonis
 C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
 C/Accession: G90603
 R/Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, I.
 Nucleic Acids Res. 29, 2145-2153, 2001
 A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pu
 A/Reference number: A99512; MUID:21267165; PMID:11353084
 A/Accession: G90603
 A/Status: preliminary
 A/Molecule type: DNA
 A/Cross-references: UNIPROT:Q98P17; GB:AL445566; PID:g14090150; PIDN:CAC13908.1; GSPDB
 A/Experimental source: strain UAB CT1P
 C/Genetics:
 A/Genetic code: SGC3

Query Match 5.4%; Score 217.5; DB 2; Length 1546;

Best Local Similarity 20.2%; Pred. No. 0.077;

Matches 180; Conservative 125; Mismatches 336; Indels 249; Gaps 40;
 Qy 1 KLGIAESKFNKLGNGKESLK-----KDTTGVEHHQENESIKESKSTFIDRNISTIRD 56
 Db 103 KSEKLASQVGSVNNBEKNFRILBAKOTKTKDNPSAKSQVLDPSQDHLKSQSKSNIN 162
 Qy 57 FENKDLKLIKKGREVDFTSRTGKMEEDYDYKD-DKGNIIAYDGDGLFETEKLD 115
 Db 163 INNEOSKQLQTLNNEBSANLQTONLLESEXTQDDLPN-----NSNKEKELKN 214
 Qy 116 IKSXYGVLSPSKDGFEILIK-INSVSNNAKYVGNNTKSIKIKATKYDFSKTWTPTDL 174
 Db 215 VLS--FNGQAVDGSFERYFYKIVSKLSEKINIDKQVDETLTDKTFADF-----SPHQ 267

QY 175 VANINDIYDGLAFAGDMRLFYQNDQKAEIKIRMEPKIKETSEKPYVSYGVNVEIGSR 234
 DB 268 Y--IETIKOLFKRAST--IKDTYO-----TNKIFLTYDEY----- 238
 QY 235 GDLSSKRPDNLFTMB-----SGKIYS-----DSKQOYLKKNIIILKRGVAL 276
 DB 299 --IKKNBSNLDKQKPSNPNFEBIYKPLEKDALVBIQALHKOYLEKNEIKITGEIYD 356
 QY 277 KTTNPGKTMLENGG-VYSKEDIKTI-----OKANPNTALIS 314
 DB 357 KVAALFKSRBELSKAGALLFSSKQOSAKITOLNHLISRYEPAPENILELSKAGQILNEL 416
 QY 315 ETTIYADSRNVEDGSGT--QSVLMSALDGFNIIRYQVTFPMQNDKGAIDKQGLHLYDSSK 373
 DB 417 KEETIYVEKGEVBIAYVTLQDAISNADQOKIFL-----NKNLKLDK--SIVVDKN- 464
 QY 374 LVLPGDKDKRYTGED--KENVBAIKEDSGMLFIDTKPVNLSMDKNYFNPB----- 421
 DB 465 ITIPAKSNVTITRKOSKSFMTPIYQKALTFEIAEPSSQSIMLNGLSGTFKQESSLVKL 524
 QY 422 -KSNKTYRANPEPYLAGKISDKGFWMLRVNBSYVDNTLIYGDHLIDNTRDPN----- 474
 DB 525 ENNAKLVAKTGTAFLNKSFSKRYG-----SVFEN--YGSVYIEGAKIWNVSESG 572
 QY 475 -----IKLNVKQGDIDMDGKQKYGKANGPDK-VTMDGNYVLTQGYSDLNAKAVGVH 525
 DB 573 GIIRNHWGSSLTPKNGEI-----RDNIISG--DKGIYISQGNALISGSDIKMS----- 620
 QY 526 YQFLYDNPVEVNIIDPKGNTSIEYADKSVENINDKR-----NNGPDGELIQEHIYING- 580
 DB 621 FSSILNLE-KTNINSGSIVNNAVSVKSLFEIDNSKIQISNNALINPGSSAIFLKN 679
 QY 581 -----KEYTSFNDIKQIIDKTLNIKIV----- 622
 DB 680 STMHLAGSLKIKKEASBQRIEVLVLPFOAKLISPENIISLDNYQGLSSAIFKTSVKN 739
 QY 603 VQDPAR-----NTTVGEFILNKDT-----GEVSELKPHRVVTVIIONGKMSSTIVSE- 652
 DB 740 INDFFKVPFLVMTYKKEFFFLMPTKLFVNPYKTLKQNHDLILOSDFESTREKIIKDELDF 799
 QY 653 -----ILPVYKSELEKGYQF-DGWEISGFEKKT-----DAGVYINLSK 689
 DB 800 YRPTAAVAKLILTLQVLTIPKHKFWEBAFEYFNPFLQKQVETINELIRIDPFYFLDIAY 859
 QY 690 DFFIKVPFKIIEKKEBENKPTFVDSKKDNPQVNSQUNESHKEDLQR 739
 DB 860 PEFV-----ENGQMLPEPPT-IHTNVNPFVLEH-----FRNEDVAR 894

RESULT 14

B42771
 reticulocyte-binding protein 2 - Plasmodium vivax (fragment)
 C/Species: Plasmodium vivax
 C/Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #ext_change 09-Sep-1997
 C/Accession: B42771
 R/Galinzi, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
 Cell 69, 1213-1226, 1992
 A/Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
 A/Reference number: A42771; M0ID:192315338; PMID:1617731
 A/Accession: B42771
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Cross-references: GB:M88098; NID:g160627; PID:g160628
 A/Experimental source: strain Belem, merozoites
 C/Genetic: RBP2
 A/Genes: RBP2

Query Match 5.4%; Score 215.5; DB 2; Length 1252;
 Best Local Similarity 20.9%; Pred. No. 0.071;
 Matches 207; Conservative 130; Mismatches 324; Indels 331; Gaps 50;

QY 2 LGIASKPK---NLGNGKEGSLKQDTTGVEHHEHSEBSIKERSFT-IDRNI----- 51
 DB 266 LSEIKYKQCKTTRISNKRQ---KQIELEKPKRBEBSNKNVNIENINRNSBOY 322
 QY 52 -----STIRDFENDLKLKKK--FREVDPTS 78
 DB 323 LKQIEDAEKQSTKVELFHHGETTISNIFSESELGVEYSQKKINVAEDIMKEIERHNS 382
 QY 79 EFKRMEEDYKTDKDNIIAYDDGDLVEYETKDEIKSIKIVGLSPSDGHEILGKI 138
 DB 383 EIQTVQKGFQENLNKLNPEPNYDMSD-ELANNKSTNAKYLIEITNLESVGN--LSEI 437
 QY 139 SNVSKAKVYVGNYSYI--EIKATYDFSKTM-----TFDYAN-INDIVQCLAPAGD 190
 DB 438 TNLKQGEKLY-SKADIMKQIAITSENTKLEKLYKQDQSNVYVNLQITF-----B 490
 QY 191 MRLFYQND-----QKAEIYIRMEKIKK-----TK-----S 218
 DB 491 RNLIVTEKRIKRLNGIDSTITINIEGALKESKGNVEIGLEKLEIEIGKRNKLVDTKKSINS 550
 QY 219 EYFVSYGVNVEIGSGSLKPKPDLITRKESKIYSDSKQOYLKKNIIILKRGVALY 278
 DB 551 TVGNFSSLFNNPDLNQYDFRNKINDYENKK--GEIYNFEGSLNKISLND--RNASENT 605
 QY 279 TTVNPKTDMLEGGVYSKEDI--AKIQANPMLPALSFTIYADSRNVEDGSGTQSVL 335
 DB 606 SDYNSAKITRLER--QKEKVLNKEBEANQYLR-----DYKATSESR----- 646
 QY 336 MSALDGFNIIRYQVTFPRANDKGAIDKQGLHLYDSSKVLVPSKDKREYTGSDPVEAI 395
 DB 647 -----FLFNMK--ESLDKINEMIKKQLTV-----NEGNGVQVQLENI 683
 QY 396 KEDSGMLFIDTKPVNS-----MDKN-----YFPSKSNKYANPEPYLAGKI 439
 DB 684 KE-----LVDEN--NLSDLIKQATGKBEIQTITSTLAKKAKITLGHVQTSKATYGIKI 736
 QY 440 SDRGFRMBLRVNESVVD-----NYLIYGDHLIDNTRDPNIKLVND-- 481
 DB 737 TP-----BLATLELGDALKTAQELKFPESKNNVLETEMKNTNELDVHKNIDQAYK 790
 QY 482 -----GDMDGKQDYKANGPDDKVTDMQGNVTL-----QTGYSDLNAKAVGVH 525
 DB 791 VALEIILASHDEIDTKQD-----SSKLEIMENQIYLVVLIQNYRKLSSIKESAVS 844
 QY 526 YQFLYDNPV-----PEVNIDPKGNTSIEYADKSVENI-----NDRKNGPDGEIOQ 574
 DB 845 VKI--GNVSKHSELSTICSDXSNDYNTIALERQTELOHLNRSFTQEKTYNDSKLE-- 900
 QY 575 HIYNGKEYTSFNDIKQIIDKTLNIKI-----VVXD-- 605
 DB 901 -----KIKTDFESLKNAL-KTLBGEVNALKASSDNHHEHVQSKSPVNPALSEIEKETD 953
 QY 606 -PARNTTVEFILNKDTGVSILKPHRVVTV--IONGKMSSTIVSEBDITLVYKGELE 662
 DB 954 IDSLNALDBLLKGGTCEVSRKYLKQVYKESIDTBLINTIERN--VAAYLAVIK 1009
 QY 663 KGYQDFGWEISGFBGQAGVYVINSKDTFIRV-----FKKIEBKEEENKPTFD-- 713
 DB 1010 KATV-----DTYQD--VLTALNHEPRTQVSNHPEPNFPMKSSSEELTKAVYDSK 1057
 QY 714 -VSKKK-----DNPQVN-----HSQUNESHK-REDIQRBEHSQ 744
 DB 1058 TIISKIKGVILIEVENTBNMTLESSAKETLEALYNELNKKYTSILNEIYQTSNEVYLOEMKS 1117
 QY 745 KSDSTQDYTA--TVVD--KNNISKSTTN 770
 DB 1118 NADKTYDVSKIFETVLDQKSNVTYQHSINN 1149

RESULT 15

A45597
 DNA-directed RNA polymerase (EC 2.7.7.6) III largest chain - malaria parasite (Plasmo
 C/Species: Plasmodium falciparum

